Naval Surface Warfare Center
Carderock Division
West Bethesda, MD 20817-5700

NSWCCD-65-TR-2007/09 May 2007

Survivability, Structures, and Materials Department

Technical Report

General Procedure for Lifetime Seaway Load Estimation (LSLE) with Examples

by

William M. Richardson

20080227051

Approved for public release; distribution is unlimited.

NSWCCD-65-TR-2007/09 General Procedure for Lifetime Seaway Load Estimation (LSLE) with Examples

DEPARTMENT OF THE NAVY

NAVAL SURFACE WARFARE CENTER, CARDEROCK DIVISION
9500 MACARTHUR BOULEVARD
WEST BETHESDA MD 20817-5700

9110
Ser 65-44
20 Dec 07

From: Commander, Naval Surface Warfare Center, Carderock Division

To: Distribution

Subj: LIFETIME SEAWAY LOAD ESTIMATION (LSLE)

Encl: (1) NSWCCD-65-TR–2007/09, *General Procedure for Lifetime Seaway Load Estimation (LSLE) with Examples*

1. Enclosure (1) provides a consistent, unified approach for U.S. Navy calculation and evaluation of proposed lifetime seaway loads. This lifetime seaway load estimation (LSLE) procedure applies to any surface ship type. The steps in the unified procedure for estimating lifetime seaway loads using model, full-scale, simulation, or analytic data are described and illustrated with examples. The unified LSLE procedure is a first principles approach, traceable with respect to assumptions and calculations. The calculations, except possibly for estimated parameter tolerance limits, can be performed on current desktop computers. The intent is to have all pertinent theory, along with examples, in one document.

2. Comments or questions may be referred to Dr. Judy A. Conley, Code 654; telephone (301) 227-1658; e-mail, Judy.Conley@navy.mil.

E. A. RASMUSSEN
By direction

Subj: LIFETIME SEAWAY LOAD ESTIMATION (LSLE)

Distribution:

COMNAVSEASYSCOM WASHINGTON DC
[SEA 05P21 (Qualley, Nappi, Sieve)]]

CNR ARLINGTON VA [ONR 331 (Barsoum, Hess)]

Attn: Mr. H. Paul Cojeen
Chief, Naval Architecture Division (G-PSE-2)
Coast Guard Headquarters
2100 Second Street, SW
Washington, DC 20593

NAVSURFWARCEN CARDEROCKDIV
BETHESDA MD [Codes 001, 3442 (TIC), 50, 65, 651, 652, 653, 654, 654 (20 copies), 655]

Defense Technical Information Center
8725 John Kingman Road Suite 0944
Fort Belvoir, VA 22060-6218

Naval Surface Warfare Center
Carderock Division
West Bethesda, MD 20817-5700

NSWCCD-65-TR–2007/09 May 2007

Survivability, Structures, and Materials Department

Technical Report

General Procedure for Lifetime Seaway Load Estimation (LSLE) with Examples

by

William M. Richardson

Enclosure (1)

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
1-May-2007	Final	-

4. TITLE AND SUBTITLE
General Procedure for Lifetime Seaway Load Estimation (LSLE) with Examples

5a. CONTRACT NUMBER

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
William M. Richardson

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER
03-1-5400-411

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) AND ADDRESS(ES)
Naval Surface Warfare Center
Carderock Division
9500 Macarthur Boulevard
West Bethesda, MD 20817-5700

8. PERFORMING ORGANIZATION REPORT NUMBER
NSWCCD-65-TR-2007/09

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Attn ONR 331
Chief of Naval Research
One Liberty Center
875 North Randolph Street, Suite 1425
Arlington, VA 22203-1995

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution is unlimited.

13. SUPPLEMENTARY NOTES

14. ABSTRACT
This report provides a consistent, unified approach for U.S. Navy calculation and evaluation of lifetime seaway load estimation (LSLE). The procedure applies to any surface ship type. The steps in the unified procedure for estimating lifetime seaway loads using model, full-scale, simulation, or analytic data are described and illustrated with examples. The unified LSLE procedure is a first principles approach, traceable with respect to assumptions and calculations. The calculations, except possibly for estimated parameter tolerance limits, can be performed on current desktop computers. The intent is to have all pertinent theory, along with examples, in one document. New methods for estimating combined loads, for estimating distribution parameter values when both low-frequency and high-frequency components are present, and estimating extreme values when combined loads are present are described.

15. SUBJECT TERMS
ship stuctures, surface ships, seaway loads

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Dr. Judy A. Conley
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	SAR	153	19b. TELEPHONE NUMBER *(include area code)* (301)-227-1658

Standard Form 298 (Rev. 8-98)
Prescribed by ANSI Std. Z39.18

CONTENTS

TABLES

FIGURES

ABSTRACT

The primary purpose of this report is to provide a consistent, unified approach for U.S. Navy calculation and evaluation of proposed lifetime seaway loads. This lifetime seaway load estimation (LSLE) procedure applies to any surface ship type. The steps in the unified procedure for estimating lifetime seaway loads using model, full-scale, simulation, or analytic data are described and illustrated with examples. The unified LSLE procedure is a first principles approach, traceable with respect to assumptions and calculations. The calculations, except possibly for estimated parameter tolerance limits, can be performed on current desktop computers. The intent is to have all pertinent theory, along with examples, in one document. Major elements of this unified LSLE approach go back several decades and were originally used for a major ship project in the 1970s. New methods for estimating combined loads, for estimating distribution parameter values when both low frequency and high frequency components are present, and estimating extreme values when combined loads are present are described.

The steps are a) develop an operational profile, b) show theory for estimating lifetime loads both with and without using order statistics, c) show models to account for errors, d) show theory useful for nonlinear parameter estimation where *our goal is to extract as much information as possible from our (usually) one sample of data*, e) show examples of statistical distribution parameter estimation using different weighted and unweighted objective functions, f) show a method for combining vertical and lateral loads with an example, g) show methods for forming, as needed, combined wave and whip distributions using various ship examples, h) determine extreme value distributions for the examples, and i) estimate the uncertainties in the extreme values.

ADMINISTRATIVE INFORMATION

The work described herein was performed by the Structures and Composites Department, Code 654, of the Survivability, Structures and Composites Directorate under work unit 03-1-5400-411.

1.0 INTRODUCTION

The primary purpose of this report is to provide a consistent, unified approach for U.S. Navy evaluation of proposed lifetime seaway loads for surface combatants using the lifetime seaway load estimation (LSLE) procedure. (This same approach can also be applied to any vehicle or platform, naval, commercial or pleasure.) Many of the steps have been known for decades. In addition to describing existing steps and methods, this report describes new extensions to previous methods (extreme value estimation of combined loads) as well as new methods for analyzing combined loads and for building distributions for combined loads.

Different marine vehicle types such as the ship of normal form, the small waterplane area ship (SWATH), the surface effect ship (SES) and hydrofoils have different physical mechanisms for obtaining support: i.e., different dominant physics. The ship of normal form and SWATHs support themselves hydrostatically, SESes support themselves aerostatically, while hydrofoils support themselves dynamically. In all cases the magnitudes of lifetime loads due to the seaway need to be estimated.[1]

We are going to describe an approach which deals with the actual physical phenomena involved and does not depend on comparison of one ship with another: i.e., it is a first principles approach, and applies to any type of vehicle. Therefore the LSLE procedure is a more fundamental approach than a rule-based approach. It should be noted that the actual lifetime loads are due to the actual lifetime operating profile, not the user specified lifetime operating profile used in the LSLE procedure.

Lifetime loads may be estimated using the results from full-scale trials, from model tests, or from simulations.

This unified LSLE procedure uses loads collected in the time domain, i.e., with a time domain sequence of loads. A frequency domain approach is not advocated for two reasons:

a) useful frequency domain approaches usually assume linearity (superposition). Large load events may be associated with nonlinear vessel responses where the linearity assumption inherent in a response amplitude operator (RAO)[2] superposition approach is not valid;

b) lifetime loads are usually associated with high sea states where infrequent and possibly nonlinear effects such as slamming are often important.

RAOs are not useful in predicting the magnitude of slamming events because the averaging process used in making the RAO causes the magnitude of the RAO values associated with slam frequencies to be significantly underestimated. The practical consequence is that the representation of the slam magnitude which actually damages the vessel in the time domain may become so small in the frequency domain that it appears insignificant. This happens because the

---

[1] Since marine vehicles are also subject to fatigue damage, a fatigue lifetime damage estimate needs to be made. For naval vessels, underwater explosive (UNDEX) loading damage needs to be considered. Fatigue (and possibly UNDEX) may have a greater effect on scantlings than the lifetime seaway load.

[2] The RAO is the magnitude of the transfer function as a function of frequency.

magnitude of its frequency contribution, which occurs over a small fraction of the record, is divided by the record length. Put another way, for a response at any frequency to be accurately represented in an RAO, it must occur continuously over the entire record.

1.1 Method of Approach

The LSLE procedure is statistical in nature. This is appropriate because the process in nature in which we are interested (getting bending moments from a seaway and then estimating lifetime loads) is *inherently statistical* because of the random nature of wave magnitudes even in a statistically stationary sea (such as operating for several hours in sea state 6). Note that a deterministic process is one which occurs with probability = 1: it is the limit when the variability of a process goes to zero.

The LSLE procedure for estimating seaway lifetime loads for all vessel types and their variants uses order statistics in the context of the "cell method". (The first U.S. Navy application is believed to be for the 3000 ton surface effect ship (3KSES) project during the 1970's.) In order to use the cell method statistical distributions for variables of interest, such as bending moments, are needed. In some cases correlated distributions of wave and whip need to be used. Part of this report will present methods and examples for estimating the parameters of these distributions, as well as, in one case, determining what distribution to use.

If we use parameter value estimation procedures which do not assign weights to the data and if we do not care, or are not concerned about how good are our parameter value estimates (what confidence we have in these estimates)[3], we can derive formulas for a lifetime load with a specified probability of exceedance without explicitly resorting to order statistics. It will be shown that the lifetime load formulas which result are identical to those derived using order statistics.

The computations for the approach described here can be (and, for the examples shown here were) carried out on current desktop computers. This means that any naval architect can apply these methods. Problems of significantly larger size may require the use of larger computer facilities such as clusters or off-site dedicated computer facilities.

The LSLE procedure is traceable. This means that at each step of the application of this unified procedure all the assumptions and the origin of numbers can be traced - there are no hidden steps. Part of this report discusses pertinent assumptions which are used in parameter estimation.

---

[3] When designing to rule, there is an inherent, but unknown confidence in our load value. This load value, along with other specifications, such as material allowable and stress allowables, in the structural criteria, result in a structure with unknown reliability with an unknown confidence.

1.2 Cell Method

A "cell" is a collection of statistically stationary events.[4] We approximate the variation of the ocean by using sets of statistically stationary events.[5] Since there is often a certain persistence in ocean conditions the assumption that conditions within a cell do not change with time is reasonable. The finer the increments in important variables such as sea state, speed, and heading the better the approximation. For example, the theory to be derived allows for an increment of sea state 5 followed by sea state 6 followed by more sea state 5. We do not assume, for example, that all of a ship's lifetime sea state 5 occurs at once.

The events in a cell are assumed to be independent and to come from a statistically stationary process. This assumption is necessary whether we estimate lifetime loads using an order statistics approach or a non-order statistics approach. The cell method will underestimate the number of lifetime cycles for fatigue loading when whipping is present since whipping cycles are highly correlated and so are not independent.

There are two kinds of cells: input cells and response cells.

Examples of input cells are:

a) running in head seas at 10 knots in state 7 seas;
b) running in quartering seas at 10 knots in state 7 seas;
c) running in head seas at 10 knots in state 6 seas;
d) running in quartering seas at 10 knots in state 6 seas, etc.

Examples of response cells for a particular operating condition such as running in head seas at 10 knots in state 7 seas are:

a) the set of vertical bending moment peaks encountered (here there are two sets: one for hog and one for sag);
b) the set of pitch amplitude peaks (again there are two sets: one of positive peaks and one of negative peaks);
c) other physical quantities of interest such as accelerations, other moments and shears, as well as other motions.

There are usually many response cells for a single input cell: at least one for each output (response) variable of interest. It is possible to have more than one response cell for the same output variable of interest. One example is to have one response cell for non-slamming bending moment responses and another response cell for slamming bending moment responses. In this case, the number of lifetime events in each cell would likely be different since the number of slams is usually much less than the number of non-slamming events.

How would we handle rare events such as rogue waves? The same way: we need a distribution for the magnitudes of the response to rogue waves. The number of lifetime events will, however, be small.

---

[4] The earliest reference found is by Edward G.U. "Bill" Band for ABS which did not explicitly use extreme values. Further references are given in his work.

[5] An example of checking for statistical stationarity is given in Richardson [1992].

In using the cell method, we need only the statistical distribution(s) associated with each response cell. The long-term distribution for each cell occurs naturally as the number of events for each cell is increased from the experimental number of events to the lifetime number of events since we are postulating a stationary process in each cell.[6] The need to obtain an explicit statistical distribution covering all the cells is eliminated since the lifetime load with a specified probability of non-exceedance (Pne) over all the cells is computed numerically. As we vary the probability of non-exceedance we numerically trace out the entire long-term distribution probability curve. To the extent that we have confidence in the distribution(s) (and their parameter values) for each cell, we have confidence in the estimated lifetime result.

A general caution: when we use a statistical distribution over a set of "n" lifetime events in a cell, we assume that the physics in the cell doesn't change. This may not be a valid assumption. For example, a large (statistically) predicted bending moment magnitude may be physically impossible of realization. In this case there is a physical limit. This possibility arises because in a formal sense, some distributions are used which go to infinity. *Physics determines what is possible, not statistics*. If we were confident in our knowledge of a physical limit, we should use statistical distributions which are limited in their range(s).

1.3 Order Statistics

Order statistics are the statistics of a sample, such as bending moment peaks, usually arranged in increasing order of magnitude. The original time order in which they were collected is lost. By giving up time order, we can we estimate lifetime loads, obtain weights to aid in improved parameter value estimation techniques, and begin to find out how "good" these estimates are statistically. The two requirements of the theory used here are:

a) the events are statistically independent of each other, and

b) the events come from a statistically stationary process.

Independence between two events means that knowing the magnitude of one event provides no information about the magnitude of the next event. For example, for a full-size vessel, the pitch amplitude after a time lapse of a tenth of second can be well predicted knowing the pitch amplitude and velocity at the start of the tenth of a second interval, and so the two events are not independent. However, predicting the following pitch peak amplitude in a random seaway is not usually very accurate, thus the two peak amplitudes are statistically independent.[7]

A statistically stationary process is one in which the statistics of the process (except for sampling variability) don't change with time. For example, the significant wave heights of two successive 20 minute wave height samples are unlikely to be the same, but if they are close to the median of the significant wave height sampling distribution for 20 minute samples, we accept that they come from the same stationary process.

---

[6] The assumption of stationarity should be checked for any data set we use. This is particularly important if we are using data collected at sea. The usual check is to see if the variability due to sample size of the input, such as waves, is within acceptable limits.

[7] Physical independence is not necessarily the same as statistical independence. For example, in sinusoidal motion, velocity is highly correlated physically with amplitude, yet the two are statistically independent over a cycle.

Formulas for order statistics will be given later and derived in Appendix A. These formulas make no assumptions as to the form of the probability distribution involved and so are more fundamental than any probability distribution which we choose to use in these formulas.

1.4 Statistical Distributions

We have two main tasks regarding statistical distributions:

a) select an appropriate distribution (Weibull, Gaussian, lognormal, etc.), and
b) estimate the parameter values for the selected distribution.

We select a distribution based on how well it fits the data. Unfortunately there is little theory to help us select which distribution to use.

The distribution of wave heights[8] in the open ocean is well approximated by a Gaussian or normal distribution. For a narrow banded Gaussian process (energy concentrated in a relatively small frequency band) the S.O. Rice [1944, 1945] distribution of amplitudes (distance to peak from the mean) has a Rayleigh distribution as its limit.[9] The paper by Jasper [1956] shows many examples of Rayleigh responses. Consequently we may consider that the Rayleigh distribution is well established theoretically and experimentally for where it is appropriate.

Bottom pressures due to slamming were shown by Ochi [1964] to follow an exponential distribution. The exponential distribution thus has some experimental support for this type of response. However, it can be shown that the exponential distribution describes a "no memory" process, meaning that the time interval between slams is random. This has been shown not to be true.[10] Having said this, larger values are apt to be uncorrelated (since they usually do not occur in sequence). Therefore, the assumption of independence in time is approximately satisfied.

The use of other distributions is supported only to the extent that they do a good job of fitting the data. Their weakness lies in the fact that usually only a relatively small amount of data exists. This problem is especially acute for large sea states since these typically result in the largest lifetime loads. The largest load is often due to slamming. Since the frequency of slamming is low, the resulting sample size is small.

As mentioned above, if we were confident in our knowledge of a physical limit, we should use statistical distributions which are limited in their range.

---

[8] "Height" is the value read at equal time intervals and is not the amplitude of the peak from the mean level nor the crest to trough distance.

[9] The narrow band assumption may be considerably relaxed without decreasing the response magnitude very much. Hence, the Rayleigh distribution is a conservative upper limit for Gaussian driven responses [Cartwright and Longuet-Higgins, 1957].

[10] This use of the exponential distribution for modeling slamming is an example of an error in the physical model formulation. An exponential distribution can be shown to have the property of no "memory" [Feller, 1968 XIII.9], i.e., events in the past have no effect on future events. This assumption is shown by experimental evidence to not be valid for slamming [Psaraftis, 1978]. Psaraftis also derives the probability distribution associated with slamming "memory", i.e., slam events are correlated in time rather than being randomly distributed in time. It is believed that extreme value theory has not yet been applied using this distribution as the parent distribution.

Weibull Distribution

Both the Rayleigh and exponential distributions mentioned above are special cases of the Weibull distribution.[11] The three parameter Weibull probability distribution is given by

$$P = 1 - e^{-\left\{\left[(x-x_0)/sf\right]^c\right\}} \tag{1-1}$$

where

P	probability that the value of the variable is x or less
x	variable such as load or response
sf	scale factor [12]
x_0	truncation value (value below which the probability is always zero)
c	exponent (also called the slope). $c = 2$ specifies a Rayleigh distribution while $c = 1$ specifies an exponential distribution. An approximation to the Gaussian or normal distribution occurs when $c \sim 3.46$.

The probability density function (pdf) for the Weibull distribution changes character at the point where the slope $c = 1$ (exponential distribution), and has some problems in the slope range 1 to 2, thus causing problems when using a maximum likelihood approach.[13]

1.5 Parameter Estimation and Parameter Covariances

The source of almost all the difficulties in estimating parameter values, and in coming to conclusions as to how valid are these estimates, is that we have but one sample of limited size. Each sample will have different data values. Put another way, we have sampling variability.

Our goal is to extract as much information as possible from our (usually) one sample of data.

The use of order statistics facilitates this goal, and is a principal reason for using order statistics.

In general, we will estimate parameter values by minimizing some criterion. We will use weighted least squares since we will often be estimating parameters for distributions including the Weibull.

In order to estimate weights resort is made to order statistics.

The distribution of parameter value estimates (one set of parameter value estimates for each sample drawn) does not have to have the same distribution as the underlying distribution from which we draw a sample. Specifically, there is no reason to assume a Gaussian or normal

---

[11] The form of the Weibull distribution was originally developed to describe a certain distribution of smallest values. We use it to describe distribution of largest values. Strictly speaking, we should call it the "naval architectural Weibull"; however, we will simply refer to it as the Weibull distribution.

[12] The characteristic value = $x_0 + sf$.

[13] See footnote 31 (Section 5).

distribution for which the distribution of many parameter estimates (the sampling distribution) is normal.

Parameter Variances and Covariances

We not only want to estimate the parameters, but would like to have some idea of how "good" these values are. The best (and very, very time consuming) way to estimate the distribution of the parameter values is do many simulations and make histograms of the resulting parameter values. From these histograms we may find, for example, the 1% and 99% tolerance limits[14] (these would cover 98% of all parameter values). In addition, we can estimate the parameter variances and covariances from the histogram values.

The smaller the parameter variances, the better (the more precise) is the estimate.

We will use a quicker (but not as precise) method to estimate the variances and covariances of the parameter values (for example, the scale factor "sf" and slope "c" of a Weibull distribution). We use a formula derived from assuming variations in the data. In order to do this we need weights for the data. Again, we are led to order statistics as a way to obtain these weights.

Once we have estimates of the parameter variances and covariances we can use them to make estimates of the variability (uncertainty) of load values in a distribution and of *lifetime load values*.

We write these variances and covariances in the form of a matrix - the parameter covariance matrix.

For example, if we estimate the scale factor "sf" and slope "c" of a Weibull distribution the *parameter covariance matrix* of these two parameters is written as

$$\mathrm{pcovm} = \begin{bmatrix} \mathrm{var}_{sf} & \mathrm{covar}_{sf,c} \\ \mathrm{covar}_{c,sf} & \mathrm{var}_{c} \end{bmatrix} \tag{1-2}$$

where

var_{sf}	variance of scale factor "sf"
var_{c}	variance of slope "c"
$\mathrm{covar}_{sf,c}$	covariance of "sf" and "c"
$\mathrm{covar}_{c,sf}$	covariance of "c" and "sf"

It may be shown that $\mathrm{covar}_{sf,c}$ always = $\mathrm{covar}_{c,sf}$. More generally the parameter covariance matrix "pcovm" is always symmetric. The covariances may be positive or negative.

[14] Tolerance limits refer to limits found from the actual distribution. Confidence limits refer to limits found from some statistic of the distribution. For example, the mean is a statistic. Percentage points of the distribution of the mean are confidence limits. The ultimate objective is to find tolerance limits.

The order of the parameters in the matrix is arbitrary.

The parameter *correlation* matrix is defined by

$$\text{pcorm} = \begin{bmatrix} \dfrac{\text{var}_{sf}}{\text{sd}_{sf}\text{sd}_{sf}} & \dfrac{\text{covar}_{sf,c}}{\text{sd}_{sf}\text{sd}_{c}} \\ \dfrac{\text{covar}_{c,sf}}{\text{sd}_{c}\text{sd}_{sf}} & \dfrac{\text{var}_{c}}{\text{sd}_{c}\text{sd}_{c}} \end{bmatrix} = \begin{bmatrix} 1 & \dfrac{\text{covar}_{sf,c}}{\text{sd}_{sf}\text{sd}_{c}} \\ \dfrac{\text{covar}_{c,sf}}{\text{sd}_{c}\text{sd}_{sf}} & 1 \end{bmatrix} \quad (1\text{-}3)$$

where

sd_{sf} standard deviation of "sf" = $\left(\text{var}_{sf}\right)^{1/2}$

sd_{c} standard deviation of "c" = $\left(\text{var}_{c}\right)^{1/2}$

The parameter correlation matrix normalizes the variances in the parameter covariance matrix to 1, thus allowing the magnitudes of the off diagonal terms relative to the normalized variances to be easily seen. It may be shown that the magnitudes of the off diagonal terms lie between -1 (negative correlation) and +1 (positive correlation). If the term = 0, then the correlation = 0 (the term represents statistical independence).

1.6 Report Topics

The remainder of this report cover the following topics:

a) section 2: development of the operational profile;
b) section 3: estimation of lifetime design load(s) not using order statistics;
c) section 4: useful formulas from order statistics;
d) section 5: estimation of statistical parameters with examples with an uncertainty example;
e) section 6: representation of combined vertical and lateral bending;
f) section 7: examples of combined wave plus whipping distributions;
g) section 8: computation of combined fits;
h) section 9: lifetime largest (extreme) values with uncertainty examples;
i) section 10: summary, conclusions, and recommendations.

The appendices cover:

a) Appendix A: order statistics;
b) Appendix B: Weibull distribution transformations;
c) Appendix C: estimate for parameter covariance matrix.

Examples showing the effect of uncertainties due to parameter values on computed values for the Weibull distribution, extreme value distributions, and of lifetime extreme value are shown.

The important topic of interpreting the parameter estimates (how "good" are the parameter estimates) is dealt with cursorily in this report.

Readers primarily interested in a topic listed below can go directly to the pertinent section since each section is, to a large extent, self contained.

a) lifetime loads: go to section 9;
b) uncertainty in lifetime loads: go to section 9.5;
c) theory for lifetime loads: go to section 3 (no order statistics) or section 4 (order statistics);
d) combined vertical and lateral loads: go to section 6;
e) methods of building distributions for combined wave and whip: go to section 7;
f) parameter estimation: go to section 5 (section 4 for origin of weights);
g) uncertainty in estimated parameters: go to section 5.10;
h) operational profiles: go to section 2.

The sections of interest for some readers may not depend on other sections, so not all sections need to be covered by all readers. Some sections build on the material of another section as shown in Table 1-1.

Table 1-1. Report Road Map

topic	**main section**	**background section(s)**
lifetime largest loads	9	2, 3 or 4, 6, 8
uncertainty in largest lifetime loads	9.5	4
operational profile	2	
combined vertical and lateral loads	6	
computation of combined wave and whip distributions	8	7
formulation of combined wave and whip distributions	7	
lifetime load estimation without using order statistics	3	
lifetime load estimation using order statistics	4	
parameter estimation	5	4
uncertainty in parameter estimation	5.10	4

1.6.1 Comments on Computational Power and Notation

Current desktop computers, now available to every engineer (circa 2005), are more powerful than the mainframe computers of a generation ago. One major consequence is that techniques known for years[15], but previously computationally prohibitive, can now be routinely applied.[16] Some of the techniques described and illustrated in this report are examples of techniques which are not new, but could be (and should be), routinely applied - especially the ones dealing with parameter uncertainty and uncertainty estimates for derived quantities such as bending moments. (Having made an estimate of parameter values using one data sample, what is the uncertainty in, say, the bending moment?)

These techniques, which we can now use routinely, often employ equations which may have a large number of terms so that we need efficient methods of dealing with (and talking about) equations having a large number of terms. This will be illustrated using an objective function (something we want to minimize) consisting of weighted sums of squares.

We can form various objective functions using sums of squares. [In the expressions below r is a residual (= data - fit), and "q" is the number of points.]

a) unweighted sum of squares

$$\mathrm{obfj} = r_1^2 + r_2^2 + r_3^2 + \ldots + r_q^2 \tag{1-2a}$$

b) weighted sum of squares

$$\mathrm{obfj} = w_1 * r_1^2 + w_2 * r_2^2 + w_3 * r_3^2 + \ldots + w_q * r_q^2 \tag{1-2b}$$

c) weights for each sum of squares and for each product such as $x_1 * x_2$:

$$\begin{aligned} \mathrm{obfj} = \; & w_{1,1} * r_1 r_1 + w_{1,2} * r_1 r_2 + w_{1,3} * r_1 r_3 + \ldots + w_{1,q} * r_1 r_q \\ & + w_{2,1} * r_2 r_1 + w_{2,2} * r_2 r_2 + w_{2,3} * r_2 r_3 + \ldots + w_{2,q} * r_2 r_q \\ & + \ldots \\ & + w_{q,1} * r_q r_1 + w_{q,2} * r_q r_2 + w_{q,3} * r_q r_3 + \ldots + w_{q,q} * r_q r_q \end{aligned} \tag{1-2c}$$

It will be seen that the number of terms increases very rapidly (~ as q^2 , so that applying eq (1-2c) to a sample of 40 points results in ~ 40^2 or 1600 terms). It is seen that for practical size problems writing out the equations is unworkable. We need a compact notation. Such a notation is matrix notation - write our equations in terms of matrices. This also lets us write each kind of variable (residual, weight, etc.) as a separate matrix. All of the above three cases can be written as

---

15 A perusal of the references will show that some of the most significant techniques are 30 or more years old. This indicates that theory directly applicable to our needs has long been available. (The references are some which the author has found useful.)

16 All of the computations needed for this report were done on a desktop computer.

$$\text{objf} = \mathbf{r}^{\mathbf{T}} * \mathbf{V} * \mathbf{r} \qquad (1\text{-}3)$$

where "T" indicates transpose.

For case a), V is a diagonal matrix with ones on the main diagonal (the identity matrix). For case b), V is again a diagonal matrix with weights w_q on the main diagonal. For case c) V is a full matrix.

The use of the notation used in eq (1-3) lets us focus on the meaning of what is going on - the interaction between kinds of variables and the transforms needed for the application of techniques - we can let computers process the individual terms.

In general, variables written in bold are vectors or matrices, while variables not written in bold are scalars. If we had 40 points then **r** has 40 rows and 1 column, $\mathbf{r}^T$ has 1 row and 40 columns, **V** has 40 rows and 40 columns, so objf = (1 x 40)*(40 x 40)*(40 x 1) = (1 x 1) - a scalar.

2.0 OPERATIONAL PROFILE

The first step in estimating lifetime loads is to find out what at-sea conditions the ship is expected to encounter, and for how long.

The operational profile ("op profile") is the collection of all input cells. Each input cell shows an operating condition such as sea state, speed, heading, vessel operating mode (such as on or off cushion for an SES), and the amount of the vessel's lifetime spent in the cell. The amount of time spent in the cell must be specified since it will be used to estimate the number of independent lifetime events needed to estimate lifetime loads.

The operational profile should be specified by the person ordering the vessel. This has yet to fully happen within the U.S. Navy and elsewhere, so the naval architect usually determines the operational profile based on the specified service (unrestricted all season open ocean North Atlantic, etc.), mission description, op tempo and the anticipated lifetime of the vessel.

2.1 Example

An example of a general ship specification using only a time and speed profile: "The ship and installed equipment shall be designed and built to provide maximum combat ready ship up-time over a 25-year life after Government acceptance, during which time the ship will be underway for 125,000 hours." Further, "The speed-time profile for underway operations shall be as follows:

1/4 speed	9 percent
1/2 speed	23 percent
3/4 speed	38 percent
Full speed	25 percent
Flank speed	5 percent

... "

In order to turn the above directive into an unrestricted open ocean operational profile we need to use actual open ocean observed sea state information. The sea state information source used here is Bales [1983]. The percentage sea state occurrences are estimated using the results of many observations over many years employing hind-casting techniques.

In the op profile Table 2-1 cols 1 and 3 come from Bales's Table 8. Col 2 is (col 3) * 125,000. Col 4 is the row sum of cols 5 through 9.

Table 2-1. Operational Profile Example

1	2	3	4	5	6	7	8	9
				<----------------		ship speed		---------------->
sea state number	desired hours	%	total actual hours	1/4 actual hours	1/2 actual hours	3/4 actual hours	full actual hours	flank actual hours
> 8	125	0.1	125	**125**	no ops	no ops	no ops	no ops
8	2123	1.7	2123	2123	no ops	no ops	no ops	no ops
7	9491	7.6	9491	2669	**6821**	no ops	no ops	no ops
6	21853	17.5	21853	2810	7180	**11863**	no ops	no ops
5	24351	19.5	24350	864	3927	9865	**9695**	no ops
4	35340	28.3	35340	1191	5414	13600	11659	**3475**
3	24600	19.7	24600	829	3769	9467	8116	2419
2	7118	5.7	7118	641	1637	2705	1779	356
0-1	0			0	0	0	0	0
actual hrs =			**125000**	11251	28749	47500	31250	6250
desired hrs =	**125000**	**100.0**		11250	28750	47500	31250	6250
speed % =				9.0	23.0	38.0	25.0	5.0
speed =				1/4	1/2	3/4	full	flank

In the above op profile, the attempt was made to satisfy both the percentage of time in a sea state and the above speed profile. Sometimes this cannot be done. Some reasons are:

a) power limit of the ship (can't make the desired speed in the specified sea state);

b) the ship would suffer significant structural damage if operated at the desired speed in the specified sea state;

c) the ride would cause injury to people or damage to equipment or cargo;

d) the ship would be overwhelmed by the sea, etc.

These concerns may be summarized by saying that the specified speed profile may be inconsistent with the probability of occurrence of sea states. In this example, a numerical solution was possible.[17]

The set of hours in a particular sea state at particular speed is called an "operational cell" or "op cell" for short. In the above op profile there are 9 sea states and 5 speeds (columns 5 through 9) for a total of 45 possible op cells. The ship is not expected to operate in sea state 0-1 (because it doesn't exist a statistically significant amount of the time: ss 0-1 range is 0 - 0.1m or 0 - 4"), nor is it expected to operate in high sea states at higher speeds. The cells in which there are no operating hours are indicated by "no ops". This leaves 26 op cells with non-zero operating hours.

[17] Most likely other solutions exist. Note, however, that the row sums, columns 5 through 9, should add to the desired number of lifetime hours in a particular sea state while the column sums should add to the desired number of lifetime hours at a specified speed.

The upper right boundary of the op profile is called the "operational envelope" ("op envelope") and is shown in bold. In this case it consists of (ss >8, 1/4), (ss 7, 1/2), (ss 6, 3/4), (ss 5, full), and (ss 4, flank). The op envelope is important because the largest loads are usually associated with the op envelope cells. [Note, however, that the (ss 8, 1/4) op cell has a factor of 2123 / 125 = 17 more time than the (ss >8, 1/4) op cell and so might result in a lifetime load greater than the (ss >8, 1/4) op cell]. Along the op envelope cells, the highest sea states op cells often result in the largest loads.

2.1.1 Comments

Is this a reasonable op profile? This depends on the ship and ship type: a normal form aircraft carrier at close to 1000' waterline length would likely encounter no problems, while a normal form cruiser at 529' waterline length might. Note the requirement that ship speed be 3/4 of full speed for 11863 hours in sea state 6 (mean H1/3 = 16.4'), and the requirement to operate at full speed for 9695 hours in sea state 5 (mean H13 = 10.7'). In this op profile, the time in each of a set of headings is not specified, so heading flexibility is still assumed to be present. For example, the op cells might be further divided into a set of op cells covering 30 or 45 degree heading sectors.

> If the op profile is unreasonable or impossible then, without further ado, the naval architect needs to return to the sponsor and keep working with the sponsor until an op profile is specified which doesn't violate the laws of physics for the ship type or cause damage or injury.

2.2 Development of Op Profiles

An op profile should consist of:

a) lifetime at-sea hours;

b) distribution of lifetime hours by speed, sea state and possibly heading.

Op profiles may be further subdivided to account for:

c) area(s) and times of operation: coastal, open ocean world wide, winter North Atlantic;

d) special operating conditions: expected to operate in ice, etc.; for coastal, special requirements such as operate in the Columbia river bar area, draft restrictions, etc.

The total lifetime at-sea hours may be divided into a collection of op profiles including conditions specific to the vessel: for example, a surface effect ship (SES) will have at least one op profile for hullbourne operation and one op profile for cushionbourne operation since the loads are usually quite different for these two operating modes..

There may be as many op profiles as necessary (coastal, open ocean, hullbourne, cushionbourne, by headings, etc.). The sum of the hours in all the op profiles sums to the total lifetime at-sea hours.

2.3 Uses for Op Profiles

The main use of the op profile for structural work is to facilitate lifetime load estimation.[18] There are two main types of load: a) "the" once in a lifetime largest load[19], and b) fatigue loads.

For once in a lifetime loads, the number of hours in a cell is converted to the number of independent events (for example, weight-buoyancy and slam induced bending moments), and then the lifetime loads are estimated using extreme value distributions (sections 3 or 4).

For fatigue loads, the distribution of the number of events in each cell is used. For example, enter S-N curves at various stress levels to form partial sums for use in lifetime fatigue estimators such as Miner's rule. (The event count due to whipping will be underestimated.)

Note that we need to have scantlings before we can do the fatigue life check (because we need stress levels), so the once in a lifetime load, along with simultaneously acting local loads, are used to get a preliminary set of scantlings.

2.4 Replacing the Ocean's Variability by Stationary Conditions

In the cell method, the variability of the actual ocean is replaced by a set of statistically stationary conditions (cells). To the extent that actual conditions at-sea persist this is a reasonable model of the ocean. We test this assumption using hindcast data [NOC, 1983].

The data we will use consist of a set of hindcasts made every six hours for a period of 20 years for the North Atlantic. The hindcasts are based on actual observations. We will use a table showing the duration (in days) for a set of significant wave heights (h1/3) at latitude 50.0° N and longitude 35.4° W. [20]

The pertinent part of the table is shown as Table 2-2a. The table entries are how many six hour hindcasts "# hc" out of 29,016 (maximum possible total is 29,200 hindcasts for 20 years) fall in a given box. For example, if the significant wave height (h1/3) ≥ 3 ft and less than 6 ft there are 50 hindcasts which fall in the zone greater than one day and ≤ 2 days. (Bottom row Table 2-2a.)

---

18 Some other uses are for propulsion power estimation (is enough power fitted to achieve the desired speed in the higher sea states); determine acceleration levels (may need a ride-control system), etc.

19 Quotation marks are used around "the" since a more accurate statement is: "What is the load which has a specified probability of <u>not</u> being exceeded over the lifetime of the ship?' For example, suppose the a load "M" has a probability of not being exceeded (Pne) of 0.9992 in cell (ss >8, 1/4) and a Pne of 0.9998 in cell (ss 8, 1/4) . Then the lifetime probability of not exceeding in both cells is given by 0.9992 * 0.9998 = 0.999. The probability of exceedance (Pe) is 1 - Pne = 1 - 0.999 = 0.001. This topic more properly belongs to the development of structural criteria.

20 The sequence number is 20 and the gridpoint subprojection is 187 - 3.

Table 2-2a. Hindcasts by h1/3 and Duration

h1/3 ft	# hc	# hc	# hc	# hc	# hc	# hc	# hc	# hc	# hc	# hc	# hc	# hc	# hc	# hc	# hc
≥ 64															
≥ 56	1														
≥ 48	3	4	1	2	1										
≥ 40	10	13	7	5	7										
≥ 34	33	22	23	22	20	2	1								
≥ 28	59	62	42	30	72	14	4	1	1						
≥ 24	69	89	84	72	106	34	9	5	5						
≥ 20	107	89	67	74	217	82	21	12	18						
≥ 16	116	105	100	98	250	128	63	19	47	7					
≥ 12	152	103	95	90	280	144	81	47	106	27	5				
≥ 9	138	100	86	87	262	151	89	67	120	66	10	5		1	
≥ 6	84	54	53	47	178	113	68	56	139	72	25	19	5	2	
≥ 3	39	16	20	18	**50**	35	21	23	78	49	25	20	5	5	2
	0.25	0.5	0.75	1	≤ 2	≤ 3	≤ 4	≤ 5	≤ 10	≤ 20	≤ 30	≤ 60	≤ 90	≤ 180	≤ 360
	Days duration of events														

The next step is to find the number of hindcast-days. See Table 2-2b. For the h1/3 ≥ 3 ft and less than 6 ft zone greater than one day and ≤ 2 days region the number of hindcast-days is 50*[(1+2)/2] = 50*1.5 = 75. (Bottom row Table 2-2b.)

Table 2-2b. Hindcast Days

h1/3 ft	hc da	hc da	hc da	hc da	hc da	hc da	hc da	hc da	hc da	hc da	hc da	hc da	hc da	hc da	hc da
≥ 64															
≥ 56	0														
≥ 48	1	2	1	2	2										
≥ 40	3	7	5	5	11										
≥ 34	8	11	17	22	30	5	4								
≥ 28	15	31	32	30	108	35	14	5	8						
≥ 24	17	45	63	72	159	85	32	23	38						
≥ 20	27	45	50	74	326	205	74	54	135						
≥ 16	29	53	75	98	375	320	221	86	353	105					
≥ 12	38	52	71	90	420	360	284	212	795	405	125				
≥ 9	35	50	65	87	393	378	312	302	900	990	250	225		135	
≥ 6	21	27	40	47	267	283	238	252	1043	1080	625	855	375	270	
≥ 3	10	8	15	18	**75**	88	74	104	585	735	625	900	375	675	540
sum	203	329	434	545	2165	1758	1250	1035	3855	3315	1625	1980	750	1080	540
	0.25	0.5	0.75	1	≤ 2	≤ 3	≤ 4	≤ 5	≤ 10	≤ 20	≤ 30	≤ 60	≤ 90	≤ 180	≤ 360
	Days duration of events														

The total number of hindcast-days is 20861 = (203 + 329 + ... + 1080 + 540). The number of hindcast years is 14.3 = (20861 / (365 days/yr * 4 hindcasts / day).

We started with 20 years of hindcasts, so we have 20 - 14.3 = 6.7 years unaccounted for. Note that Table 2-2b ignores h1/3 of less than 3 ft. When we make a plot of the data from Bales [1983] Table 6: Annual Sea State Occurrences in the Open Ocean North Atlantic we find that 18.5% of the occurrences are less than 3 ft. Adding 0.185 * 20 years = 3.7 years to the 14.3 years already found we have a total of 18.0 years.

We have accounted for 18/20 = 90% of the occurrences using duration information. The use of the statistically stationary assumption appears justified.

2.5 Summary

The op profile summarizes the derived operating conditions and the time spent in each operating condition. Op profiles derived from observations at-sea may differ considerably from implied Gen Specs requirements and may in some cases be inconsistent with Gen Specs requirements.

Using op profiles at an early design stage facilitates the identification of crucial operating conditions for the vessel. This allows early consultation with the sponsor if potential "show stoppers" are found.

After agreement with the sponsor, the use of op profiles allows effort to be concentrated on those conditions (usually very few) which most likely drive the design.

The op profile forms a useful basis for estimating lifetime and preliminary fatigue loads.

3.0 ESTIMATION OF LIFETIME DESIGN LOAD NOT USING ORDER STATISTICS

We now know how many response events occur for each operating condition (section 2). We will use this knowledge to help derive an expression for lifetime load values with a specified probability of non exceedance (Pne) over the at-sea lifetime of the ship. The probability of exceedance (Pe) is given by Pe = 1 - Pne.

We wish to estimate a lifetime design load having a specified probability of non-exceedance, such as 0.999, over the lifetime of the ship.

Once we have an operational profile, the first piece of information needed is how many *independent* events, or responses, occur in each op cell. The number of lifetime response encounters in an op cell is found from the number of lifetime hours in the op cell.

The number of independent events may be estimated using the wave spectrum modal period, the wave spectrum average wave period, or, if frequency response spectrum information is available, by using the average response period. A better method is to use the auto-correlation function to estimate an independence interval in which the largest value is selected (an example showing the validity of this approach is shown in section 6.3). The best method is to read one value of interest, such as one hog and one sag, per pitch cycle.

If there are two process in an op cell, for example slamming and non-slamming events, the response cell may be divided into two response cells, one for slam events and one for non-slam events. These two response cells will likely have greatly different number of events since the frequency of slamming is very different from the non-slam frequency.

The second piece of information needed for each response cell is the statistical distribution of load values in the response cell.[21] The parameters of the distribution are estimated using data derived from experiments, simulations, or analytic information.[22] However the data are obtained, the resulting statistical distribution is called the parent distribution. The parent distributions for the "m" cells $cell_1$, $cell_2$, ... , $cell_{cm}$ are denoted by P_{c1}, P_{c2}, ... , P_{cm}.[23] Note that the distribution type (Weibull, etc.) for one response cell may be different than the distribution type for another response cell. Even for the same distribution type the parameter values usually vary from cell to cell.

In summary, the operational profile consists of "m" op cells with each op cell having a (usually) different number of lifetime events. Each op cell generates one or more response cells.

---

21 A commonly used distribution is the Weibull (which has the exponential and Rayleigh distributions as special cases. See eq (1-1) Section 1).

22 We will not discuss here how hull size and shape impact loads, and why those considerations often lead to using test results rather than using simulation or analytical results. Suffice to note that the LSLE method is applicable to primary loads (and some secondary loads as well) for any ship type: normal form ship, SES, SWATH, catamaran, etc.

23 By convention capital letters such as "P" denote probabilities, while small letters such as "p" denote probability densities.

Each response cell has a different parent distribution P giving the probability distribution of the magnitude of events in the cell.

The probability distribution and the number of occurrences in a response cell are the two core pieces of information needed to estimate both large (lifetime) loads and fatigue loads.

3.1 Theory Not Using Order Statistics

3.1.1 The First Response Cell

Consider the first response cell, $cell_{c1}$, having n_{c1} events. Select a load magnitude L. The probability that the load magnitude L is not exceeded is $P_{c1}(x \le L)$.

The probability that the magnitude of the first load event in $cell_{c1}$ does not exceed L is $P_{c1}(x \le L)$. The probability that the magnitude of the second load event in $cell_{c1}$ does not exceed L is also $P_{c1}(x \le L)$. The probability that both of the first two load events do not exceed L is given by the probability that event 1 does not exceed L and the probability that event 2 does not exceed L i.e., the joint probability, is given by

$$P_{c1}(x \le L) * P_{c1}(x \le L) = P_{c1}(x \le L)^2 \tag{3-1}$$

The reason why we get this simple expression is that *the events are independent*, meaning that the conditional probability of occurrence of load event 2 given that load event 1 has occurred = zero. The practical meaning of independence in this situation is that, given the magnitude of load event 1, we cannot predict the magnitude of load event 2.

We continue: the probability that the magnitude of the third load event in $cell_{c1}$ does not exceed L is also $P_{c1}(x \le L)$. The probability that the magnitude of the fourth load event in $cell_{c1}$ does not exceed L is also $P_{c1}(x \le L)$. The probability that all of the first four load events do not exceed L is given by

$$P_{c1}(x \le L) * P_{c1}(x \le L) * P_{c1}(x \le L) * P_{c1}(x \le L) = P_{c1}(x \le L)^4 \tag{3-2}$$

We trudge on through all the n_{c1} events in response $cell_{c1}$: the probability that all of the n_{c1} load events do not exceed L is given by the product of n_{c1} multiplies of $P_{c1}(x \le L)$:

$$[n_{c1} \text{ terms}]: \; P_{c1}(x \le L) * P_{c1}(x \le L) * \ldots * P_{c1}(x \le L) * P_{c1}(x \le L) = P_{c1}(x \le L)^{nc1} \tag{3-3a}$$

which is written as

$$F_{c1}(x \le L) = P_{c1}(x \le L)^{nc1} \tag{3-3b}$$

$F_{c1}(x \le L)$ is the probability that none of the n_{c1} load events in response $cell_{c1}$ exceeded the load magnitude L. $F_{c1}(x \le L)$ is the probability of non-exceedance Pne for response $cell_{c1}$. It will be shown in section 4 dealing with order statistics that this formula is the extreme value distribution for response cell c1 found using order statistics.

Being able to relate our expression for "F" to order statistics means that the extensive theory associated with order statistics becomes available when it comes time to estimate distribution parameter values and their covariances. This leads to the ability to establish confidence values for the distribution parameter values and allows uncertainty estimates to be made. An important hidden assumption in "L" As a Design Load (section 3.2) is that for some reason we knew the distribution parameter values exactly.

3.1.2 The Second Response Cell

Consider the second response cell, cell$_{c2}$, having n_{c2} events. Select the same load magnitude L as for response cell$_{c1}$. The probability that the load magnitude L is not exceeded is $P_{c2}(x \leq L)$. Recall that P_{c2} is not necessarily the same as P_{c1}.

Go through the same procedure as for response cell$_{c1}$.

The probability that all of the n_{c2} load events do not exceed L is given by the product of n_{c2} multiplies of $P_{c2}(x \leq L)$:

$$[n_{c2} \text{ terms}]: \; P_{c2}(x \leq L) * P_{c2}(x \leq L) * \ldots * P_{c2}(x \leq L) * P_{c2}(x \leq L) = P_{c2}(x \leq L)^{nc2} \qquad (3\text{-}4a)$$

which is written as

$$F_{c2}(x \leq L) = P_{c2}(x \leq L)^{nc2} \qquad (3\text{-}4b)$$

$F_{c2}(x \leq L)$ is the probability that none of the n_{c2} load events in response cell$_{c2}$ exceeded the load magnitude L. $F_{c2}(x \leq L)$ is the probability of non-exceedance Pne for response cell$_{c2}$. It will be shown in section 4, dealing with order statistics, that this formula is the extreme value distribution for response cell c2 found using order statistics.

3.1.3 Probability of Not Exceeding Load Magnitude L in the First Two Response Cells

The probability of not exceeding the load magnitude L in the first two response cells is given by

$$F_{1,2}(L) = [P_{c1}(x \leq L)^{nc1}] * [P_{c2}(x \leq L)^{nc2}] = F_{c1}(x \leq L) * F_{c2}(x \leq L) \qquad (3\text{-}5)$$

In words: the probability that none of the n_{c1} and the n_{c2} load events are greater than load magnitude L is given by $F_{1,2}(L)$.

The reason we get this relatively simple expression for $F_{1,2}(L)$ is that the response cells are independent since we cannot simultaneously operate in response cell$_{c1}$ and in response cell$_{c2}$.

3.1.4 Probability of Not Exceeding Load Magnitude L in Any Response Cell

The same load magnitude L is used for all response cells. The probability of not exceeding the load magnitude L in any of the "m" response cells is given by

$$F(L) = P_{c1}(x \leq L)^{nc1} * P_{c2}(x \leq L)^{nc2} * \ldots * P_{cm}(x \leq L)^{ncm} \qquad [\text{m terms}] \qquad (3\text{-}6a)$$

$$= F_{c1}(x \leq L) * F_{c2}(x \leq L) * \ldots * F_{cm}(x \leq L) \qquad [\text{m terms}] \qquad (3\text{-}6b)$$

In words: the probability that <u>none</u> of the $n_{c1} + n_{c2} + ... + n_{cm} = n$ lifetime load events are greater than load magnitude L is given by F(L). Here "n" is the lifetime number of events.

F(L) is often written as Pne(L), and is the <u>lifetime probability of non-exceedance</u> of the responses to the entire operational profile for load L.

3.1.5 Probability of Exceedance "Pe"

The probability of exceedance for the load value L common to all cells is given by

$$Pe(L) = 1 - Pne(L) \tag{3-7}$$

This is the probability of exceedance "Pe" (which we usually wish to make small). One interpretation of the load L is that if the load L is exceeded, we are willing to live with the consequences.

An example using four response cells is shown in Table 3-1. The lifetime hours are <u>not</u> those shown in Table 2-1, but are representative. We wish to estimate the probability of exceedance of a load value = 256,000 ft-lt = 256 k-ft-lt. A Weibull distribution, eq (1-1), is assumed for all four response cells.

Columns 1 and 2 are for identification: their values do not enter the computations. Column 5 = Integer[(col 3)*3600/(col 4) + 0.5] is the number of lifetime events for this cell. The lifetime number of independent events (556100) is the last entry in column 5, and equals the sum of the number of independent events for each cell.

Columns 6, 7, 8 are the parameters for the probability distribution for each cell (in this case, Weibull, though any probability distribution can be used for any cell).

Column 9 is the probability corresponding to a load of 256 k-ft-lt in each cell.[24] This is the probability that a single event will have a magnitude > 256 k-ft-lt. For cell >8, v1, we have

$$P = 1 - e^{-\left\{[(x-x0)/sf]^{c}\right\}} = 1 - e^{-\left\{[(256-0)/54]^{1.95}\right\}} = 0.999\,999\,999.$$

Column 10 is the probability that none of the loads in a cell exceed a value of 256 k-ft-lt. For example, in cell (>8, v1), 0.999 996 640 is the probability that none of the 3,600 encounters in cell (>8, v1) will exceed 256 k-ft-lt. For cell >8, v1 we have

$$F = P^{n} = 0.999\,999\,999^{3600} = 0.999\,996\,640.$$

Column 11 contains the probability calculation that none of the events in any cell exceed 256 k-ft-lt, and is obtained by multiplying together the probabilities of <u>not</u> exceeding in any cell. For example, 0.999 004 020 = 0.999 996 640 * 0.999 007 376. The overall Pne = 0.999 996 640 * 0.999 007 376 * 1.000 000 000 * 0.999 999 999 = 0.999 004 019.

---

24 The calculations are actually done in double precision. Probability values truncated to 9 significant figures are shown in Table 3-1.

Table 3-1. Probability of Exceedance of a 256 k-ft-lt Load

1	2	3	4	5	6	7	8
cell id	ss	lifetime hours	av enc period, sec	lifetime encoun	<- parent prob distribution x_0	sf	c ->
>8, v1	> 8	10	10	3600	0	54	1.95
8, v1	8	200	9	80000	0	60	2.00
7, v1	7	250	8	112500	0	42	2.05
7, v2	7	700	7	360000	0	48	2.10
...							
lifetime number of independent encounters =				556100			

Table 3-1 continued

1	2	9	10	11	12	13
cell id	ss	P eq (1-1)	$F = P^n$	Fne calculation $= \prod F_k$	1 - F	1 - P
>8, v1	> 8	0.999 999 999	0.999 996 640	0.999 996 640	3.36E-06	9.33E-10
8, v1	8	0.999 999 988	0.999 007 376	0.999 004 020	9.93E-04	1.24E-08
7, v1	7	1.000 000 000	1.000 000 000	0.999 004 020	under flo	under flo
7, v2	7	1.000 000 000	0.999 999 999	0.999 004 019	8.79E-10	2.49E-15
...						
lifetime probability of non exceedance (Pne) =				0.999 004 019		
lifetime probability of exceedance (Pe) = 1 - Pne				0.000 995 981		

The value in col 11 for the last cell (7, v2) is the probability we are looking for: Pne = 0.999 004 019 is the probability that none of the 556100 independent encounters exceeded 256,000 ft-lt. The probability of exceedance over the lifetime of the ship is 1 - Pne = 0.000 995 981 or about 0.001.

The values in columns 12 and 13 are interpreted as follows:[25] the probability in col 12 is the probability of exceedance in a cell of a load value 256 k-ft-lt for any of the loads in the cell, while the probability in col 13 is the probability that a single load exceeds 256 k-ft-lt.

[25] "under flo" indicates that we have lost enough precision so that none of figures are significant. If the calculations are arranged as shown, the values of Pne and Pe are valid unless Pe is on the order of $< 10^{-9}$ or smaller in which case an investigation into loss of precision should be made. Few things in life have a probability of exceeding of 10^{-9} (one in a billion), so this is rarely of concern.

It is <u>not</u> correct to sum the 1 - F values in col 12 to obtain the overall Pe.[26]

The values in column 12 enable a quick determination of the critical cells. It is seen that cell (8, v1) is the main contributor to Pe since its contribution is about 300 times that of the next largest contributor.

<u>Note that these Pne calculations can be (and were) done on a spreadsheet</u>.

3.2 "L" As a Design Load

The comments in sections 3.2.1, 3.2.2, and 3.2.3 refer to the structural capacity or resistance to the load "L". L is a candidate for the structural design load in the following senses:

3.2.1 No Variability in Structural Parameters

If material properties were exact, if analysis methods gave exact values for strength (such as buckling), if construction had no errors (exact fits, etc.), and all other factors affecting the structural resistance of the ship were known exactly (no uncertainty), then the factor of ignorance[27] associated with the structural resistance = 1.0.

Designing to load L with a factor of ignorance = 1.0 would produce a ship having a lifetime reliability = Pne.

3.2.2 Variability in Structural Parameters

Since, in real life, the factor of ignorance is > 1.0, the probability distributions of as many factors (materials, structural assemblies, etc.) affecting structural resistance as can be found are used to form the joint probability distribution of load and structural resistance over the lifetime of the ship. We desire that the probability of not exceeding the joint probability of load and resistance have a specified value such as 0.999 (the analog to our load Pne).

Note that when using this approach we no longer have "a" design load since the structural resistance is now a probability distribution, not a single value. Instead, we have a distribution of loads each of which has a certain probability of not being exceeded by the structural resistance. The joint distribution of load and structural resistance may be thought of as providing a weighting factor for any particular combination of load and structural resistance. We are interested in the final result: the probability that the joint probability of load and structural resistance achieve a minimum specified value such as 0.999 over the life of the ship. An approach for handling this real life situation where both loads and structural resistance (strength) are described by statistical distributions is shown in Richardson [1996]. A computer program (ProbE, by the author) for

---

26 To see this, imagine that we have 200 cells, each with 1-F = 0.01. Summing the 1-F values results in 200*0.01 = 2: a nonsensical value, since, by definition, probability values lie between 0 and 1. The correct Pne for this case is $0.99^{200} = 0.134$, so Pe = 1-0.134 = 0.866.

27 The "factor of safety" is composed of two parts: a true factor of safety (for example, a submarine collapse depth is greater than the operating depth) and a factor of ignorance which covers uncertainties in design, manufacturing and use.

estimating the joint probability of non exceedance Pne for a ship section using a level III reliability approach has been written.

To get started in design, we specify a load L_1 such that the joint probability equals, say, 0.999. L_1 will be greater than L since L_1 has to account for uncertainties in material, construction, and analysis methods.

3.2.3 Approximations for Obtaining Joint Structural Pne

Suppose that we do not have the necessary material and structural distributions? There are a variety of approaches using the specific load "L". Their usefulness depends upon the fact that the variability of the other factors influencing structural reliability are in some sense small compared with the variability in the loads produced by the seaway.

Some examples are:

a) set a factor of ignorance for all factors on the structure, and proceed;

b1) find statistical distributions for material properties, and define a minimum mechanical property value such as requiring that 99% of all material values, such as yield strength exceed this value;

b2) define a minimum mechanical property such as yield strength, and 100% test all material, accepting only that material meeting the minimum;

c) set a factor of safety, based on experiments, on values such as buckling strength; and require that this value be used;

d) use b1 or b2 in conjunction with c;

e) use ultimates.

Note that the combination of the load "L" used and the factor(s) used can be adjusted so long as the combined effect results in the desired joint Pne.

In all cases the effect of uncertainty in the structural analysis methods used must be taken into account.

3.3 Effect of Varying the Common Load L

Each individual cell has a well defined extreme value distribution $F = P^n$. Its pdf is given by $pdf(F) = n * P^{n-1} * p$. It can be shown that, for large sample sizes (number of encounters), pdf(F) has a maximum at F = 1/e = 0.368. This means that 1 - 1/e = 0.632 (63.2%) of all largest values are greater than the most probable extreme value.

The dominant cell in Table 3-1 is cell (8, v1) and has a Rayleigh distribution (c = 2). Is 80,000 encounters a large sample size in the sense of the above paragraph? The probability density peak for cell (8, v1) occurs at F = 0.384467, whereas 1/e = 0.367879. The difference in probability is 0.0165878. The difference in x values is 202.0 (x for F=0.384467) - 201.6 (x for F=0.367879) = 0.4 k-ft-lt. The 1/e approximation results in a difference of 0.4 / 202 = ~ 0.2% difference in the load.

To obtain the probability density function (pdf) of the largest values using all the cells we use the following procedure. Vary the load L to obtain $F(L_1), F(L_2), \ldots, F(L_k)$ for "k" different values of the common load L (which is the same for all response cells). If these are plotted the cumulative distribution of the lifetime non-exceedance distribution is obtained. The probability density function (pdf) of this distribution may be approximated by forming

$$pdf(F_{k,k-1}) = [\ F(L_k) - F(L_{k-1})\ ] \ /\ (\ L_k - L_{k-1}\ ) \tag{3-8}$$

The probability density peak occurs at F = 0.384231, whereas 1/e = 0.367879. The difference in probability is 0.0163511. The difference in x values is 202.1 - 201.7 = 0.4 k-ft-lt. The 1/e approximation results in a difference of 0.4 / 202 = ~ 0.2% difference. The reason that the values are about the same as for the dominant cell is that the dominant cell contributes (Table 3-1 continued) 0.999 004 019 / 0.999 007 376 ~ 100% of the non-exceedance.

A plot of $pdf(F_{k,k-1})$ is shown in Figure 3-1. The lifetime extreme value probability density distribution is non-symmetric with the longer tail to the right.

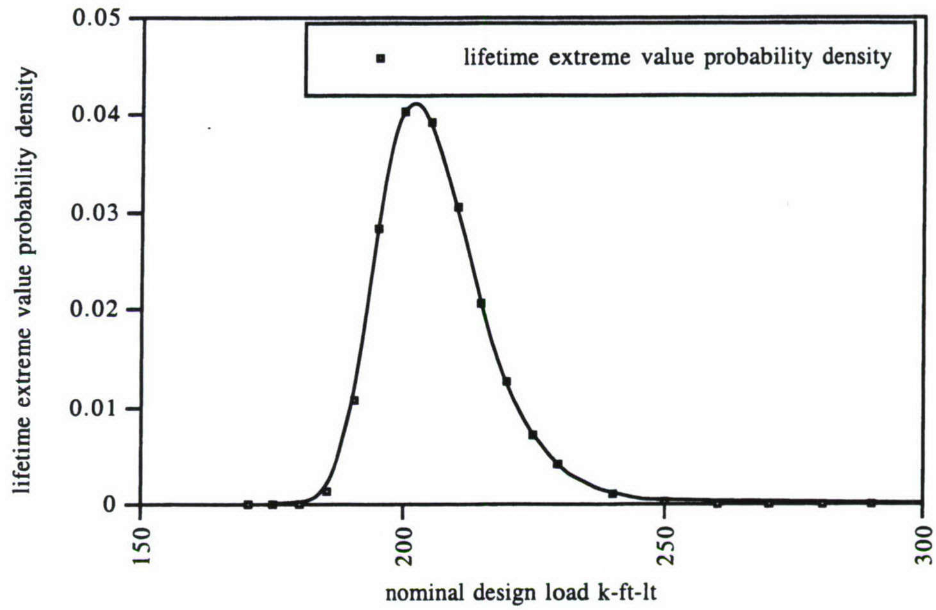


Figure 3-1. Op Profile Lifetime Extreme Value Probability Density

Figure 3-2 shows the load associated with lifetime probabilities of exceedance Pe of 0.9, 0.632, 0.1, 0.05, 0.01, 0.001, and 0.0001. The value associated with Pne = 1/e =0.368 (so Pe = 1 - Pne = 0.632) is close to the most probable extreme load. It is exceeded by 0.632 = 63% of all extreme values and so is <u>not</u> a suitable candidate for a design load.

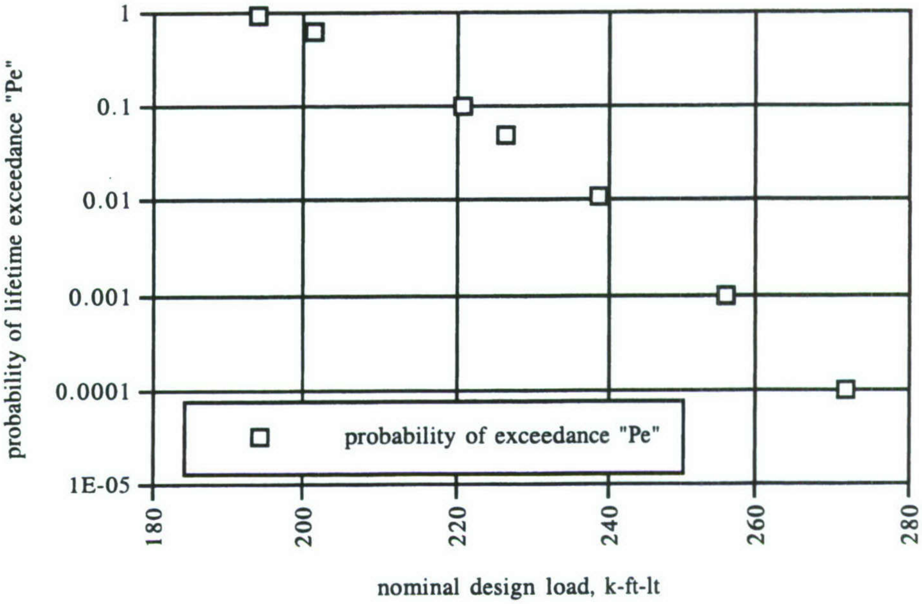


Figure 3-2. Probability of Exceedance "Pe" vs. Load

A plot of load ratio vs. Pe ratio is shown in Figure 3-3. The base values are load = 256 k-ft-lt and Pe = 0.001 over the life of the ship so that both the load ratio and the Pe ratio = 1 for these values (upper left-hand corner). The chance of encountering a load 1% larger than 256 k-ft-lt (= 259 k-ft-lt) over the lifetime of the ship is about 0.001*0.69 = 0.00069 = 0.07%. The chance of encountering a load 5% larger than 256 k-ft-lt (= 269 k-ft-lt) over the lifetime of the ship is about 0.001*0.15 = 0.00015 = 0.02%. The chance of encountering a load 10% larger than 256 k-ft-lt (= 282 k-ft-lt) over the lifetime of the ship is about 0.001*0.022 = 0.000022 = 0.002%.

Since material properties are often quoted as a minimum such that 99% of all material samples have properties greater than the minimum, there is usually a strength margin in the improbable case that a larger load than the base load of 256 k-ft-lt is encountered. Of greater concern is that buckling margins may be less than the material margins, thus making this the critical condition.

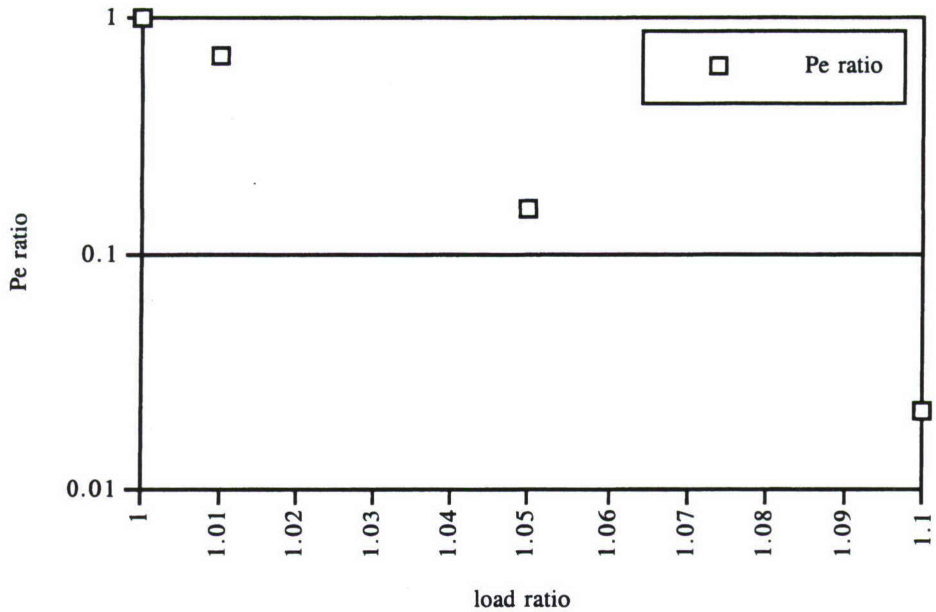


Figure 3-3. Load Ratio vs. Pe Ratio

3.4 Summary

A method has been shown whereby a load with a specified probability of exceedance "Pe" over the lifetime of the ship can be found. Theory not involving order statistics has been developed and applied to the operational profile responses to obtain this load.

When there is a small probability of exceedance, plots have been developed to show that the sensitivity of Pe to loads greater than the nominal design load is small.

Methods on how to use this load for design depending upon the degree of uncertainty in material properties, construction variability, and analysis methods have been sketched out.

4.0 ORDER STATISTICS

Since we have:

a) formulas which allow us to find the relation between load and the probability of non-exceedance Pne (section 3), and

b) methods for estimating distribution parameter values which do not require order statistics (section 5), why do we need the complication of order statistics?

There are at least four reasons for doing so:

a) The link between the method previously used to obtain the lifetime load distribution and order statistics will become clearer.

b) When doing parameter estimation we will find that better practice is to use weights for the members of a sample set. These weights can be obtained using order statistics.

c) After obtaining parameter estimates, we may want to know how good these are. Order statistics, in conjunction with formulas to be displayed later, will enable us to approximate the parameter covariance matrix without resorting to simulation.

d) We can obtain relatively simply an expression giving the probability that both the largest and the second largest load in a cell will exceed a specified value. This will increase our confidence that it is sufficient to use only the extreme value distribution for lifetime load estimation (example not shown in this report).

Order statistics are the statistics of a sample, such as bending moment peaks, usually arranged in increasing order of magnitude. The two requirements are:

a) the events are statistically independent of each other, and

b) the events come from a statistically stationary process.

The two most important order statistics are those of the largest in a sample of loads (such as a sample of bending moments) and the smallest in a sample of structural resistances (such as a sample of stiffeners). Since our interest is in loads, our primary focus will be on the largest in a sample.

4.1 Distribution of the r^{th} Value in a Sample Size n

The formula for the probability distribution of the r^{th} observation in a sample n is a function only of the parent distribution and the number of encounters or events.

The formula (derived in Appendix A) for the probability density of the r^{th} observation in a sample of size n, where $r = 1$ is the smallest magnitude observed and $r = n$ is the largest, or extreme magnitude observed is

$$f_r \quad = \quad \frac{n!}{(n-r)!\,(r-1)!} \quad * P^{(r-1)} \quad * \quad (1-P)^{(n-r)} \quad * \quad p \qquad (4\text{-}1)$$

where

f_r probability density distribution for r^{th} observation

r rank order by magnitude ($r = 1$ is smallest, $r = n$ is largest)

n sample size
P parent probability distribution
p parent probability density distribution

This expression is fundamental in the sense that it applies to any probability distribution. As mentioned above, it depends only upon the parent distribution "P" and the sample size "n".

For example, the expression for the probability density of the second largest value in a sample is obtained by setting $r = n-1$ to obtain

$$f_{n-1} = n*(n-1) * P^{(n-2)} * (1-P) * p \tag{4-2}$$

To find the probability F_{n-1} we must resort to numerical integration.

The probability density of the smallest value is obtained by setting $r = 1$ to obtain

$$f_1 = n * (1-P)^{(n-1)} * p \tag{4-3a}$$

This may be directly integrated to obtain

$$F_1 = 1 - (1-P)^n \tag{4-3b}$$

as the probability distribution of the smallest value in a sample of size n.

4.2 Extreme Values

The extreme, or largest value distribution probability density equation is obtained by setting $r = n$ in eq (4-1) to obtain

$$f_n = n * P^{(n-1)} * p \tag{4-4a}$$

The above expression may be directly integrated to give the probability $F(n)$ associated with an extreme value

$$F_n = P^n \tag{4-4b}$$

Note that this is the same formula as was obtained before (section 3 eqs 3-3b and 3-4b).

If we know the largest in a sample, why are we going to need a distribution of largest values? Why is not the largest the largest? The answer to these questions results from considering sampling variability.

Suppose we go down the test tank once and record 200 peak bending moments. One of these will be the largest in the first sample. Suppose we go down the test tank a second time and again record 200 peak bending moments. One of these will be the largest in the second sample, and is quite unlikely to have the same value as the largest value in the first sample of 200.

Suppose we go down the test tank a third time and again record 200 peak bending moments. One of these will be the largest in the third sample, and is quite unlikely to have the same value as the largest value in either the first or second sample of 200. Suppose we have unlimited funds and go down the tank 1,000,000 times. We will obtain the 1,000,000 largest values from our 1,000,000 samples of size 200. We can form a histogram of these. This histogram forms an approximation to the extreme, or largest value distribution for sample size 200 for peak bending moments.

4.2.1 Weibull Distribution Example

The Weibull distribution probability given by eq (1-1) is repeated and renumbered

$$P = 1 - e^{-\left\{\left[(x-x_0)/sf\right]^c\right\}} \tag{4-5a}$$

where

P	probability that the value of the variable is x or less
x	variable such as load or response
sf	scale factor [28]
x_0	truncation value (value below which the probability is always zero)
c	exponent (also called the "slope")

$c = 2$ specifies a Rayleigh distribution which characterizes a weight-buoyancy (W-B) distribution as well as narrow banded motion responses to a Gaussian (normal) input,[29] while $c = 1$ specifies an exponential distribution which characterizes a slamming pressure distribution.

The probability density "p" for the Weibull distribution is given by

$$p = \frac{\partial P}{\partial x} = \left(\frac{c}{sf}\right) * \left\{\left[\frac{(x-x_0)}{sf}\right]^{(c-1)}\right\} * e^{-\left\{\left[(x-x_0)/sf\right]^c\right\}} \tag{4-5b}$$

The mean of the Weibull distribution is given by

$$\text{mean}_{\text{Weibull}} = x_0 + sf * \Gamma\left(1+\frac{1}{c}\right) \tag{4-5c}$$

The variance of the Weibull distribution is given by

$$\text{var}_{\text{Weibull}} = sf^2 * \left[\Gamma\left(1+\frac{2}{c}\right) - \Gamma^2\left(1+\frac{1}{c}\right)\right] \tag{4-5d}$$

where

Γ gamma function [= generalized factorial of (m+1), m = 0, ... , m]
$\Gamma(0+1) = 0! = 1,\ \Gamma(1+1) = 1! = 1,\ \Gamma(2+1) = 2! = 2,\ \Gamma(3+1) = 3! = 6$, etc.

---

28 The characteristic value = $x_0 + sf$.

29 Open ocean wave heights have been shown to be well represented by a Gaussian distribution.

Plots of the parent distribution probability density, extreme value probability density, and the extreme value probability for two Weibull distributions are shown on the next page (Figure 4-1). The left hand side shows the parent probability distribution for an exponential distribution ($c = 1$), its extreme or largest value probability density function for 100 events, and the extreme value probability for the 100 events. The right hand side shows the parent probability distribution for a Rayleigh distribution ($c = 2$), its extreme or largest value probability density function for 100,000 events, and the extreme value probability for 100,000 events. For the particular distribution parameter values shown the exponential distribution dominates even though there are only 0.1% as many events.

The conclusion is that distribution type, distribution parameter values, and number of events all need to be considered in determining the most severe condition.

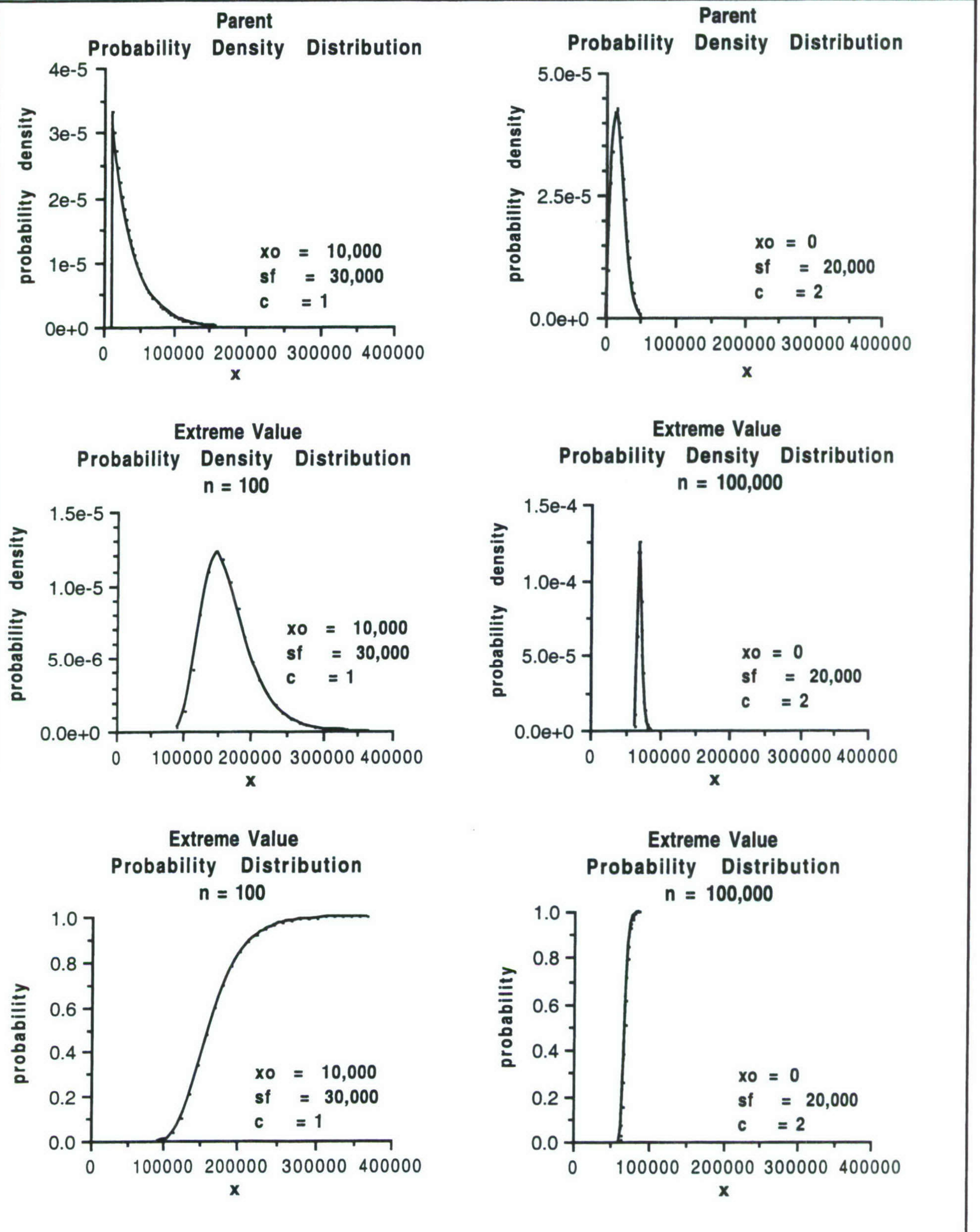


Figure 4-1. Parent and Extreme Value Distributions

4.3 Link between Experiment and Extreme Values

"P" is also referred to as the experimental distribution for the following reason. Suppose we obtain a sample of 200 values by doing an experiment. We would then use these 200 values to estimate parameter values for the experimental distribution. This allows us to directly link experimental results to how they are used in the extreme value distribution since we see that the parent distribution does not change with sample size (because we have assumed statistical stationarity). The distribution $F_{200} = P^{200}$ gives the distribution of extreme values obtained by repeatedly obtaining samples of 200, while $F_{10000} = P^{10000}$ gives the distribution of largest, or extreme values for a lifetime of 10,000 events.

4.4 Probability of Not Exceeding Load Magnitude L in Any Response Cell

The <u>same</u> load magnitude L is used for all response cells. The probability of <u>not</u> exceeding the load magnitude L in any of the "m" response cells is given by

$$F(L) = P_{c1}(x \le L)^{nc1} * P_{c2}(x \le L)^{nc2} * \ldots * P_{cm}(x \le L)^{ncm} \quad \text{[m terms]} \qquad (4\text{-}6a)$$
$$= F_{c1}(x \le L) * F_{c2}(x \le L) * \ldots * F_{cm}(x \le L) \quad \text{[m terms]} \qquad (4\text{-}6b)$$

In words: the probability that <u>none</u> of the $n_{c1} + n_{c2} + \ldots + n_{cm} = n$ lifetime load events are greater than load magnitude L is given by F(L). Here "n" is the lifetime number of events.

F(L) is often written as Pne(L), and is the <u>lifetime probability of non-exceedance</u> of the entire operational profile for load L.

A compact notation for F(L) is $$F(L) = Pne(L) = \prod_{k=1}^{m} [P_k(L)]^{n_k} \qquad (4\text{-}6c)$$

4.5 Order Statistics Mean, Variance, Covariance

When doing parameter estimation taking into account the variability associated with each point we need expressions for the variance of a single order statistic "r" and the covariance of two order statistics "r" and "s". The covariance requires an expression for the joint distribution of two order statistics.

We are <u>not</u> speaking here of the variance of the entire distribution such as eq (4-5d), but of the variance associated with a particular order number.

The expressions for the order statistics variances and covariances (Appendix A) are

$$\sigma_{r:n}^2 = \int_{-\infty}^{\infty} (x - \mu_{r:n})^2 * f_r(x) * dx \qquad (4\text{-}7a)$$

and

$$\sigma_{rs:n} = \int_{-\infty}^{\infty} \int_{-\infty}^{y} (x - \mu_{r:n}) * (y - \mu_{s:n}) * f_{rs}(x, y) * dx * dy \qquad (4\text{-}7b)$$

It may be shown that $\operatorname{cov}(x,y) = \operatorname{cov}(y,x)$. In other words, the covariance matrix is always symmetric.

The variances and covariances may be written as a covariance matrix:

$$\sigma_{rs:n} = \begin{bmatrix} \sigma^2_{1:n} & \sigma_{1,2:n} & \cdots & \sigma_{1,n-1:n} & \sigma_{1,n:n} \\ \sigma_{2,1:n} & \sigma^2_{2:n} & \cdots & \sigma_{2,n-1:n} & \sigma_{2,n:n} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \sigma_{n-1,1:n} & \sigma_{n-1,2:n} & \cdots & \sigma_{n-1,n-1:n} & \sigma_{n-1,n:n} \\ \sigma_{n,1:n} & \sigma_{n,2:n} & \cdots & \sigma_{n,n-1:n} & \sigma^2_{n:n} \end{bmatrix} \tag{4-7c}$$

The above expressions involve the means and the probability density functions of a single order statistic and of two order statistics. The means are given by

$$\mu_{r:n} \equiv E(x_{r:n}) = \int_{-\infty}^{\infty} x * f_r(x) * dx \tag{4-8a}$$

where

E expected value operator

The means may be written as a column vector

$$\mu_{r:n} \equiv \begin{bmatrix} \int_{-\infty}^{\infty} x * f_1(x) * dx \\ \int_{-\infty}^{\infty} x * f_2(x) * dx \\ \cdots \\ \int_{-\infty}^{\infty} x * f_{n-1}(x) * dx \\ \int_{-\infty}^{\infty} x * f_n(x) * dx \end{bmatrix} \tag{4-8b}$$

The probability density function of a single order statistic has been given above:

$$f_r = \frac{n!}{(n-r)!\,(r-1)!} * P^{(r-1)} * (1-P)^{(n-r)} * p \tag{4-1}$$

The joint probability density of two order statistics is given by

$$f_{rs}(x,y) = K_{rs}\, P^{r-1}(x)\, p(x)\, [P(y)-P(x)]^{s-r-1}\, p(y)\, [1-P(y)]^{n-s} \tag{4-9}$$

where

$$K_{rs} = \frac{n!}{(r-1)!\,(s-r-1)!\,(n-s)!}$$

When considering two order numbers, "s" refers to the second order number and can range from r+1 through n. "r" is always less than "s".

Here "y" is a dummy variable of integration: the distributions "P(x)" and "P(y)" are the same distribution.

We now have the expressions needed for parameter estimation with weights and for estimation of the parameter covariance matrix.

4.5.1 Reduced Weibull

The Weibull distribution may be written in reduced form. Define $y = \frac{x - x0}{sf}$. Then the Weibull distribution becomes

$$P = 1.0 - e^{-\left\{y^c\right\}} \tag{4-10a}$$

with probability density

$$p = \frac{\partial P}{\partial y} = c * \left\{y^{(c-1)}\right\} * e^{-\left\{y^c\right\}} \tag{4-10b}$$

The mean of the distribution in original space is

$$x_{mean} = x_0 + sf * y_{mean} \tag{4-11a}$$

The variance of the distribution in original space is

$$x_{var} = \left(sf^2\right) * y_{var} \tag{4-11b}$$

Doing computations in reduced space means dealing with only one variable, the slope "c", instead of with three (x_0, sf, and c). The results in original space are obtained using eqs (4-11).

4.5.2 Order Statistics Variance Examples

The variance associated with an order number should not be confused with the variance of the entire distribution such as eq (4-5d).

Each order point (r = 1, 2, ... , n) has its own mean and variance. In Table 4-1 for c = 1 (exponential) (col 1) and c = 2 (Rayleigh) (col 2), the variance of the largest or extreme value is largest. (The spread of the pdf of the extreme is greater than that of any other order point.)

The progression of increasing variance with order number does not have to hold for all distributions, or even for all parameter values for a specific distribution.. It is possible for variances to start large, decrease, and then become larger again. An example for the Weibull distribution is shown in Table 4-1 for sample size n = 9 with c = 10.

For $c = 10$ (col 4), the variance of the smallest point $r = 1$ is the largest variance. The variance for $r = 7$ is the smallest. The corresponding standard deviations are also shown. The columns for $c = 1$ and $c = 2$ correspond to the plots in Figure 4-2.

Table 4-1. Reduced Weibull Distribution Variances for c = 1, 2 and 10, n = 9

1	2	3	4	5	6	7
	var	var	var	st dev	st dev	st dev
r	c = 1	c = 2	c = 10	c = 1	c = 2	c = 10
1	0.01235	0.02384	0.00844	0.11111	0.15442	0.09188
2	0.02797	0.02755	0.00436	0.16724	0.16598	0.06602
3	0.04838	0.03040	0.00306	0.21995	0.17436	0.05533
4	0.07616	0.03357	0.00243	0.27596	0.18322	0.04931
5	0.11616	0.03762	0.00207	0.34082	0.19395	0.04554
6	0.17866	0.04333	0.00186	0.42268	0.20815	0.04318
7	0.28977	0.05242	0.00177	0.53830	0.22895	0.04204
8	0.53977	0.07006	0.00181	0.73469	0.26469	0.04250
9	1.53977	0.12433	0.00222	1.24087	0.35260	0.04711

The variance of the order points is also a function of sample size. The variances of the largest points are shown in Table 4-2 for sample sizes 9, 18, and 27. Whether the variance of the largest order point increases or decreases with sample size depends upon the slope.

Table 4-2. Effect of Sample Size on Largest Value Variance

1	2	3	4	5	6
	var	var	st dev	st dev	probability covered
r = n	c = 1	c = 2	c = 1	c = 2	= 1/(n+1)
9	1.53977	0.12433	1.24087	0.35260	0.10000
18	1.59089	0.10430	1.26131	0.32296	0.05263
27	1.60857	0.09492	1.26830	0.30809	0.03571

The change in variance with increasing sample size is relatively gradual, and does not necessarily become smaller with increasing sample size [Table 4-2 (col 2)]. By contrast, the corresponding change in the amount of probability covered by a given order number is relatively smaller, and converges to 0 as the sample size becomes large (col 6). These observations have relevance when we come to do parameter estimation for the Weibull distribution [section 5].

In reduced space ($x_0 = 0$, $sf = 1$) for $c = 1$ (exponential) the variance of the entire Weibull distribution = 1, while for $c = 2$ (Rayleigh) the variance = 0.21460 (st dev = 0.46325).

Figures 4-2 are plots of Weibull distribution probability density functions in reduced space for the exponential (Figure 4-2a) and the Rayleigh (Figure 4-2b) distributions for $n = 9$.

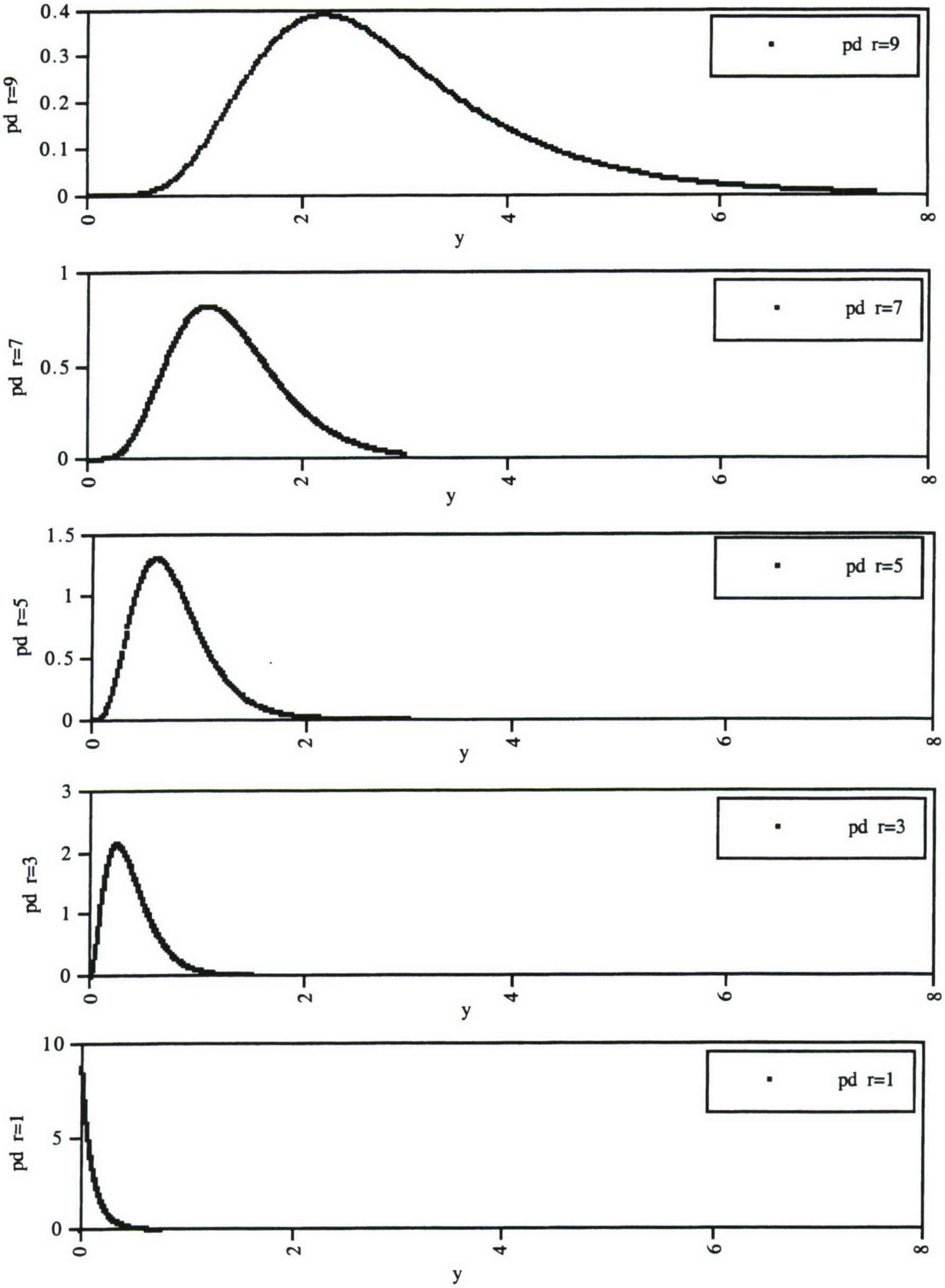


Figure 4-2a. Weibull c=1, Probability Density Plots for r=1, 3, 5, 7, 9, n = 9
(Vertical Axes Scales Differ)

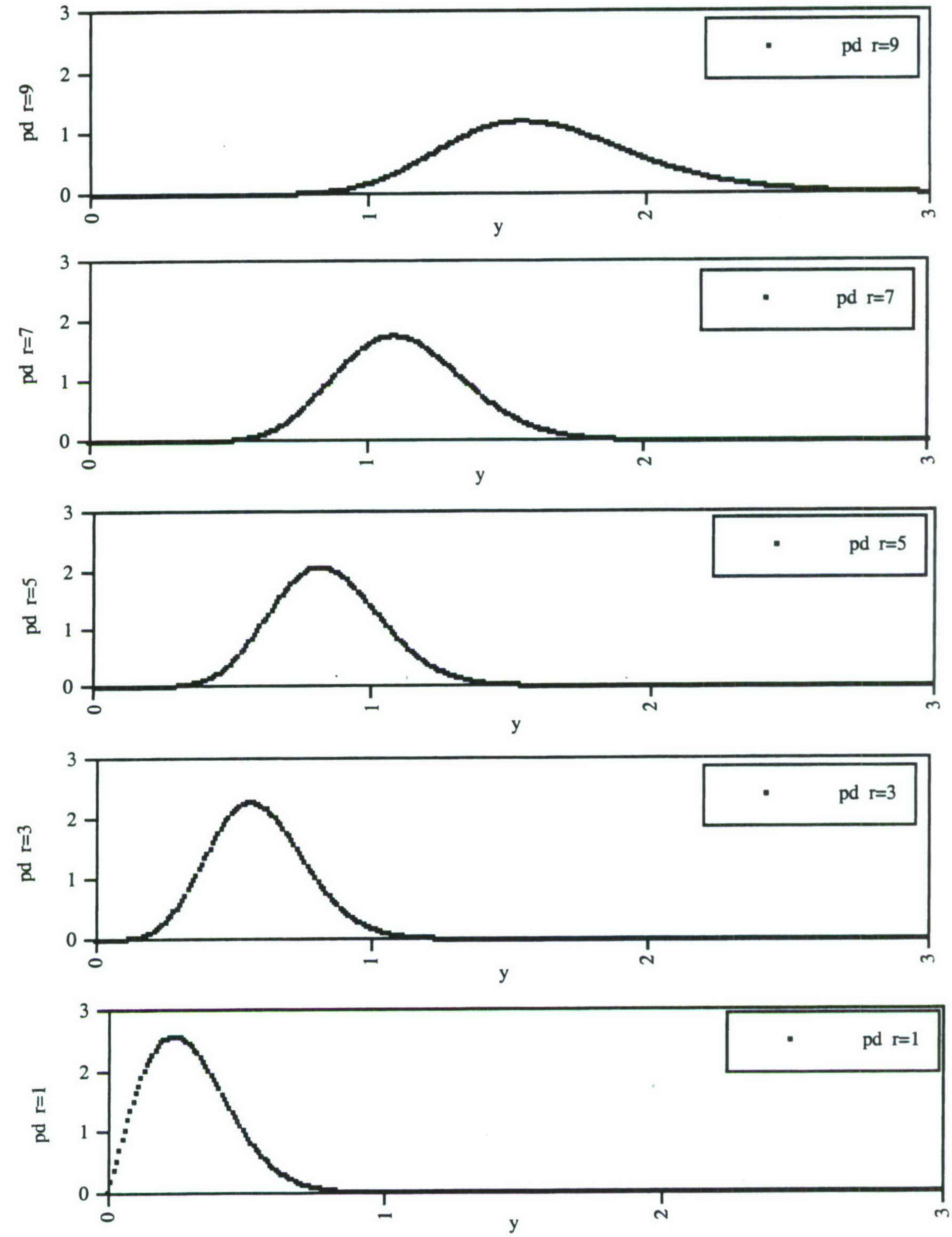


Figure 4-2b. Weibull c=2, Probability Density Plots for r=1, 3, 5, 7, 9, n = 9
(Vertical Axes Scale Same)

5.0 PARAMETER AND PARAMETER COVARIANCE ESTIMATION

We are concerned with estimating parameter values for distributions which we feel describe the data. We wish not only to estimate parameter values, but to form some idea as to how good are these estimates.

By a distribution we mean a function which we feel is a plausible candidate to fit the observed data. The distribution may consist of a single distribution with fixed parameter values (such as a Weibull distribution with fixed "sf" and "c"), a single distribution with varying parameter values (such as a Weibull distribution with "sf" and "c" given as functions of the values of another distribution), or more than one distribution (one distribution for low-frequency components and another distribution for high-frequency components).

Our distributions are non-linear in the parameters to be estimated [sometimes linear after transformations (mappings) such as into Weibull space[30]], so we are engaged in nonlinear parameter estimation. Some of the methods employed here are significantly influenced by the discussions and derivations in Bard [1974].

The usual way to estimate parameter values is to minimize some function of the data. Candidate functions are:

a) maximum likelihood,
b) the sum of squares of the residuals (residual = data - fit),
c) weighted sum of squares of the residuals.

We will use weighted least squares since we will often be estimating parameters for a Weibull distribution. The pdf for the Weibull distribution changes character at the point where the slope c = 1 (exponential distribution), thus causing problems when using a maximum likelihood approach.[31]

Before we show examples of parameter estimation (section 5.8) and an estimate for parameter variability (section 5.9) it is useful to describe the conceptual framework within which parameter estimation is done (sections 5.1 - 5.5). *This conceptual framework applies to any fitting function where we estimate parameter values no matter whether the fitting function is a statistical distribution or not.*

---

30 See Appendix B for the transformation (mapping) into Weibull space.

31 The report author's poor experience with the maximum likelihood method is consistent with Castillo [1988, §5.3.1.2] which states, " The [maximum likelihood] system [of equations] ... must be solved iteratively, with the help of a computer. However, we need to point out here that, in some cases, a solution cannot be found, i.e., a relative maximum does not exist. It is also worthwhile mentioning that the Weibull pdf presents a discontinuity at [c] = 1, and for [c] < 1 and [c] > 1 this family of distributions shows different behavior at the lower finite end. For [c] < 1, the likelihood function can be made infinity if [x_0] is made equal to the minimum of the sample values. For this reason, and for the reason that some regularity problems occur for 1 < [c] < 2, the maximum likelihood method is not appropriate for Weibull populations with values of [c] < 2." The principal range of interest for ship loading problems covers the range 1 < [c] < 2. While the range of "c" values of interest to Weibull is not known, and no examples are given, his report [Weibull, 1967] was done for the Air Force in the context of lifetime testing. A later report by Abernethy et. al. [1983] which also seems mainly concerned with failures, shows most examples with slopes "c" > 2 (and some with c > 9). Weibull goes on to state, regarding using the maximum likelihood method, "... and, above all, the estimates from small samples are heavily biased."

The following subsections discuss or show:
a) types of uncertainties and errors (section 5.1);
b) sampling distributions (section 5.2);
c) forms of error models (section 5.3);
d) objective function formulation (section 5.4);
e) weighting (section 5.5);
f) assignment of probability (section 5.6);
g) objective function example (section 5.7);
h) parameter estimation using Weibull distribution examples (section 5.8);
i) parameter covariance estimation (a start on how good are the parameter estimates) with examples (section 5.9).
j) comments on "goodness" of fit, confidence intervals, tolerance intervals and limits (section 5.10).

The methods described below can be applied to any distribution. Methods specific to Weibull space (section 5.8.1, 5.8.2, and 5.8.3) would be replaced by methods appropriate to other distributions or else not be used.

5.1 Uncertainties and Errors

It is customary to refer to differences between observed values and their underlying true values as errors. This viewpoint works well when the variability of the physical process being sampled is small compared to the errors of observation (such as when doing a calibration). The implication is that with better and better measurements we could reduce the errors of observation to some arbitrarily small value: i.e., errors are something we have control over.

This is not a useful way to think when a random process having an inherently large variability is being sampled. A better term is uncertainty, since variability is inherent in our random process (random seaway driven responses), and so is not something over which we have control no matter what the sample size.[32] (The variance of, say, a Rayleigh distribution is a property of the distribution, not of the sample size of the samples we take.)

We will use the term "uncertainty" rather than "error" in the remainder of this report to remind us that a significant amount of the variability in what we deal with has to do with inherent variability, not with variability which we can control. This more general viewpoint will allow us to formulate a unified approach for dealing with uncertainty whether or not we can control it.

We have (mathematical) models with parameter values to be estimated. Five sources of uncertainty in the resulting parameter values are due to:
a) measurement uncertainty;
b) sample size effects;
c) the use of different models;
d) different estimation procedures, and
e) assignment of probability if we are dealing with statistical distributions (section 5.6).

[32] Often by increasing the sample size we can, however, reduce the variability in estimated parameter values.

For our situation (working with towing tank data) measurement uncertainty is composed of instrumentation uncertainty and calibration uncertainty.[33] These effects are small compared to the other effects, and will be set = zero. In case they are not insignificant their variance may be added to the variances associated with sample size uncertainty since there is unlikely to be any correlation between the value measured and the size of the instrumentation and calibration uncertainty.

Sample size variability is due to drawing a finite size sample from a large population. Repeated samples of, say, size nine will have a distribution for each order number: if we drew 2000 samples of size nine, we would have 2000 smallest values ($r = 1$), 2000 next to smallest values ($r = 2$), ... , 2000 next to largest values ($r = n-1$), and 2000 largest values ($r = n$).

Suppose we had a very large sample size precisely measured so that the variability in the estimated parameter values is very small. The estimated parameter values would still likely vary from (math) model to (math) model: the parameter values depend, in general, upon the model. As discussed above, model refers to the distribution that we expect the data to follow such as Weibull, Gaussian, or some other.

In general, we use some estimation procedure applied to a data sample of a specified size. Estimation procedures are not restricted to formulas for means, variances, etc., but can be thought of as any algorithm used to produce estimates of a statistic[34] or a parameter value. Consequently, different estimation methods applied to the same (math) model with the same data can produce different parameter values. For example, for a nonlinear problem we would not expect parameter values estimated using maximum likelihood to be the same as when using weighted least squares.[35]

5.2 Sampling Distributions

We will use the term "sampling distribution" in a slightly wider context than is usually employed.

The usual context is to speak of the distribution of a statistic of repeated samples of size "n". For example, we draw repeated samples of size nine, and compute the mean, variance, mode, etc. of these samples. We form histograms of these statistics. These histograms are an approximation to the sampling distribution of the distribution mean, variance, mode, etc. for sample size nine. These distributions have their own means, variances, etc. These estimation procedures have no unknown parameters; they are computed from the data sample.

We can extend this idea to distributions of estimated parameter values formed by repeatedly drawing samples of a given size (such as of size nine) and estimating distribution parameters. We form histograms of these parameter values. These histograms are an approximation to the sampling distribution of the parameter values for sample size nine. These

---

[33] Theory has been developed to deal with the case where the independent variables have uncertainty. The resulting models are more complex and almost always nonlinear.

[34] A statistic is something computed or estimated from a sample such as the mean, variance, mode, etc.

[35] For linear models, the results are often the same.

distributions have their own means, variances, etc. In addition, we may form covariance estimates where it makes sense: for example, between different parameters.

Define an estimate[36]

$$\boldsymbol{\phi}^* = \mathbf{h}(\mathbf{W}) \tag{5-1}$$

where

$\boldsymbol{\phi}^*$	point estimate of a parameter vector
$\mathbf{h}$	estimation procedure (data mean, variance, parameter estimation, etc.)
$\mathbf{W}$	data matrix (of finite size) comprising both dependent and independent variables

In general, there is a different sampling distribution of $\boldsymbol{\phi}^*$ for each sample size, each mathematical model (fitting function), and each estimation procedure.

$\boldsymbol{\phi}^*$ has a sampling distribution which depends upon both the estimation procedure and the data matrix. The data matrix depends upon the underlying distribution of all the data $\boldsymbol{\Omega}$. The distribution of $\boldsymbol{\Omega}$ depends upon the true values $\hat{\boldsymbol{\phi}}$ which are usually unknown (else we would not be trying to estimate them).

5.2.1 Mean Square Error

The expected value of $\boldsymbol{\phi}^*$ may have both systematic error (bias) and random error. A true measure of how good is the estimate is given by the mean square error

$$E\left(\boldsymbol{\phi}^* - \hat{\boldsymbol{\phi}}\right)\left(\boldsymbol{\phi}^* - \hat{\boldsymbol{\phi}}\right)^{\mathbf{T}} = E\left(\boldsymbol{\phi}^* - \bar{\boldsymbol{\phi}}\right)\left(\boldsymbol{\phi}^* - \bar{\boldsymbol{\phi}}\right)^{\mathbf{T}} + \left(\bar{\boldsymbol{\phi}} - \hat{\boldsymbol{\phi}}\right)\left(\bar{\boldsymbol{\phi}} - \hat{\boldsymbol{\phi}}\right)^{\mathbf{T}} = \mathbf{V}_{\boldsymbol{\phi}} + \mathbf{b} * \mathbf{b}^{\mathbf{T}} \tag{5-2}$$

where

E	expected value operator
$\bar{\boldsymbol{\phi}}$	mean (expected) value of the estimate $\boldsymbol{\phi}^*$
$\mathbf{V}_{\boldsymbol{\phi}}$	variance of the estimate $\boldsymbol{\phi}^*$
$\mathbf{b}$	bias of the estimate $\boldsymbol{\phi}^* = \bar{\boldsymbol{\phi}} - \hat{\boldsymbol{\phi}}$
$\hat{\boldsymbol{\phi}}$	true value of the estimate
$\mathbf{T}$	transpose

This expression is valid whether we are dealing with a vector of estimators such as mean and variance, a vector of parameters such as a Weibull "sf" and "c", or a vector of observations such as the r^{th} in a sample of "n".

To compute the expectation directly we need to know two sampling distributions. First, we need the sampling distribution of true values so that we may compute $\hat{\boldsymbol{\phi}}$. Second, we need the sampling distribution of expected values so that we may compute $\bar{\boldsymbol{\phi}}$. Recall that $\bar{\boldsymbol{\phi}}$ depends

36 The notation in Bard [1974] is extensively used in the remainder of this section.

upon the estimator (the estimating procedure), so that the result is valid for only one estimating procedure. Alternatively, we could, using simulation, estimate the variance $\mathbf{V}_{\boldsymbol{\phi}}$ and the bias $\bar{\boldsymbol{\phi}} - \hat{\boldsymbol{\phi}}$. This still implies that we know the sampling distribution of true values $\hat{\boldsymbol{\phi}}$.

We would like to minimize the mean square error. There are several reasons why we will not be able to do so:

a) We do not have enough knowledge to find the sampling distribution(s). (If we did we would not be trying to estimate - we could most likely compute values of interest.) This is in contrast with the Gaussian or normal distribution where results concerning the sampling distribution of means, variances, ratios of variances, etc. are known.

b) The parameter estimation procedures we will use may have bias. In the absence of simulations we do not have an estimate of the bias.

c) We used some procedure to estimate the parameter values. In general, different estimation procedures applied to the same data will yield different parameter values.

d) We have only one sample with which to work. If we use simulation to estimate the covariance matrix of the parameters, we use as the assumed correct values the parameter values we estimated using our one set of data with a particular estimation procedure.

5.3 Form of Error Models

The errors in the observations are given by

$$\boldsymbol{\varepsilon}_{\mu} \equiv \mathbf{w}_{\mu} - \hat{\mathbf{w}}_{\mu} \tag{5-3}$$

where $\boldsymbol{\varepsilon}_{\mu}$ error(s) for the μ^{th} observation, $\mu = 1,2,...,n$

$\mathbf{w}_{\mu}$ measured value(s) for the μ^{th} observation

$\hat{\mathbf{w}}_{\mu}$ true value(s) of the μ^{th} observation

This equation applies to both independent and dependent variables.

For example, we obtain a sample of "n" observations. These might be a set of nine maximum bending moments obtained during a ship model test run in the towing tank.

Suppose we have a fitting model where we can solve for the dependent variable(s) in terms of the independent variable(s).[37] If there is no measurement or sample size error such a fitting model may be written in the form [38]

$$\hat{\mathbf{y}}_{\mu} = \mathbf{f}\left(\hat{\mathbf{x}}_{\mu}, \boldsymbol{\theta}\right) + \boldsymbol{\varepsilon}_{\mu 2} \tag{5-4}$$

[37] A more general form of a model occurs when we cannot explicitly solve for the dependent variable. While statistical techniques can still be applied to estimated parameter values, the difficulties increase.

[38] In this section, we modify and add to the formulation in Bard [1974] in order to formulate equations suitable for our needs.

where

$\mathbf{f}$	set of fitting model equations (= 1 for single equation least squares)
$\hat{\mathbf{y}}_\mu$	true value of the dependent variables(s) for the μ^{th} observation, $\mu = 1,2,\ldots,n$
$\hat{\mathbf{x}}_\mu$	true value of the independent variable(s) for the μ^{th} observation
$\boldsymbol{\theta}$	parameters whose values we are trying to estimate
$\boldsymbol{\varepsilon}_{\mu 2}$	error in the fitting model equations

A residual is defined as [observed - (fitting model)] where the value of "fitting model" is obtained using the estimated value of the parameters. If we used the true (unknown) values of the parameters $\boldsymbol{\theta}$ we would still have a residual since the errors $\boldsymbol{\varepsilon}_{\mu 2}$ in the fitting model equations are still present. If the fitting model equations were an exact representation of the physical situation ($\boldsymbol{\varepsilon}_{\mu 2} = 0$), then the residuals = errors in the fitting model equations.

5.3.1 Uncertainties in Both Independent and Dependent Variables

Now suppose there are uncertainties and errors in both the independent and dependent variables. Define residuals for both the independent and dependent variables:

$$\mathbf{e}_{\mathbf{x}\mu}\left(\hat{\mathbf{x}}_\mu\right) \equiv \mathbf{x}_\mu - \hat{\mathbf{x}}_\mu \tag{5-5a}$$

and

$$\mathbf{e}_{\mathbf{y}\mu}\left(\boldsymbol{\theta}, \hat{\mathbf{x}}_\mu\right) \equiv \mathbf{y}_\mu - \left(\mathbf{f}\left(\hat{\mathbf{x}}_\mu, \boldsymbol{\theta}\right) + \boldsymbol{\varepsilon}_\mu\right) \tag{5-5b}$$

where

$\mathbf{x}_\mu$	observed value of the independent variable(s) in the μ^{th} observation
$\mathbf{y}_\mu$	observed value of the dependent variable(s) in the μ^{th} observation
$\boldsymbol{\varepsilon}_\mu \equiv \boldsymbol{\varepsilon}_{\mu 1} + \boldsymbol{\varepsilon}_{\mu 2} + \boldsymbol{\varepsilon}_{\mu 3}$	
$\boldsymbol{\varepsilon}_{\mu 1}$	uncertainty in $\mathbf{y}_\mu$ (due to measurement and calibration error)
$\boldsymbol{\varepsilon}_{\mu 2}$	error in the fitting model equations
$\boldsymbol{\varepsilon}_{\mu 3}$	uncertainty in $\mathbf{y}_\mu$ due to sampling variability

Make a single vector of residuals

$$\mathbf{e}_\mu\left(\boldsymbol{\theta}, \hat{\mathbf{x}}_\mu\right) \equiv \begin{pmatrix} \mathbf{e}_{\mathbf{x}\mu} \\ \mathbf{e}_{\mathbf{y}\mu} \end{pmatrix} \tag{5-6}$$

We will form some function of the residuals eq (5-6) in order to estimate values of the unknowns ($\boldsymbol{\theta}$ and $\hat{\mathbf{x}}_\mu$). Eq (5-6) is the most general form of the error model we will consider.

If at all possible, we wish to avoid the situation of unknown values of the independent variables since the more unknown values we need to find, the greater the uncertainty in those values. The total amount of information in a given sample is constant: the more information we use in estimating the values of unknowns, the less information is available for estimating uncertainty. *If we have as many unknowns as observations, then we use all the information to estimate the values of the unknowns, and so the uncertainty in the estimated values is infinite.*

5.3.2 Uncertainty Exists Only in the Sample Observations

If the measurement and calibration errors are sufficiently small compared to the sampling variability they may be neglected. In this case eq (5-5a) vanishes and eq (5-6) simplifies. We obtain

$$\mathbf{e}_{\boldsymbol{\mu}}(\boldsymbol{\theta}) = \mathbf{y}_{\boldsymbol{\mu}} - \left(\mathbf{f}\left(\hat{\mathbf{x}}_{\boldsymbol{\mu}}, \boldsymbol{\theta}\right) + \boldsymbol{\varepsilon}_{\boldsymbol{\mu}23}\right) \tag{5-7}$$

where

$\boldsymbol{\varepsilon}_{\boldsymbol{\mu}23} \equiv \boldsymbol{\varepsilon}_{\boldsymbol{\mu}2} + \boldsymbol{\varepsilon}_{\boldsymbol{\mu}3}$ (measurement and calibration error neglected)

We will form some function of the residuals eq (5-7) which we will minimize in order to estimate values of the parameter unknowns $\boldsymbol{\theta}$ in eq (5-7).

The error in the model equations $\boldsymbol{\varepsilon}_{\boldsymbol{\mu}2}$ still is present as is the uncertainty $\boldsymbol{\varepsilon}_{\boldsymbol{\mu}3}$ in $\mathbf{y}_{\boldsymbol{\mu}}$ due to sampling variability. We will be estimating parameter values with these sources of uncertainty present. Eq (5-7) is the equation we will use for the examples.

If, after using a large sample size to estimate parameter values, we are left with large values of the residuals, then it is an indication that our model is likely wretched.

5.4 Objective Function

Why do we need an objective function? The problem arises because eq (5-6) or eq (5-7) can be satisfied by an infinite number of sets of values of $\boldsymbol{\theta}$ and $\hat{\mathbf{x}}_{\boldsymbol{\mu}}$ [eq (5-6)] or by any set of values of $\boldsymbol{\theta}$ [eq (5-7)]. (We have yet to choose a fitting function.) We wish to choose one set of these values. We do this by minimizing some function of the residuals eq (5-6) if we need to estimate both $\boldsymbol{\theta}$ and $\hat{\mathbf{x}}_{\boldsymbol{\mu}}$, or by minimizing some function of the residuals eq (5-7) if we need only to estimate $\boldsymbol{\theta}$. This function is called the objective function.

Since we are going to use least squares, form the weighted sum of squares

$$\text{objf} = \mathbf{e}_{\boldsymbol{\mu}}\left(\boldsymbol{\theta}, \hat{\mathbf{x}}_{\boldsymbol{\mu}}\right)^{\mathbf{T}} * \mathbf{V} * \mathbf{e}_{\boldsymbol{\mu}}\left(\boldsymbol{\theta}, \hat{\mathbf{x}}_{\boldsymbol{\mu}}\right) \tag{5-8a}$$

or

$$\text{objf} = \mathbf{e}_{\boldsymbol{\mu}}(\boldsymbol{\theta})^{\mathbf{T}} * \mathbf{V} * \mathbf{e}_{\boldsymbol{\mu}}(\boldsymbol{\theta}) \tag{5-8b}$$

where

objf objective function (a scalar)
$\mathbf{e}_{\boldsymbol{\mu}}$ vector of residuals

V weighting matrix

We wish to estimate values of the parameters $\boldsymbol{\theta}$ and the true values of the independent variables $\hat{\mathbf{x}}_{\mu}$ which minimize the value of eq (5-8a). If we only need values of the parameters $\boldsymbol{\theta}$ we minimize the value of eq (5-8b).

In a formal sense the solution is given by solving the following set of equations:[39]

$$\begin{aligned} \partial \mathrm{objf}/\partial z_1 &= 0 \\ \partial \mathrm{objf}/\partial z_2 &= 0 \\ &\cdots \\ \partial \mathrm{objf}/\partial z_m &= 0 \end{aligned} \tag{5-9}$$

where

z_1 , z2	first, second unknown parameters, etc.
m	number of unknown parameters

In general, we need a nonlinear equation solver to find the parameter values. This is not a great problem since some spreadsheets have nonlinear equation solvers, or, if such a spreadsheet is not available, computer programs are available to do this task.

For a Weibull distribution with $x_0 = 0$, using eq (5-8b), we wish to estimate two parameters: "sf" and "c".

We draw a sample of "n" observations. This might be the set of maximum bending moments obtained during a model run in the towing tank.

For the examples to be shown, we have one equation which we will apply to each of "n" observations. The equation has one independent variable and one dependent variable.

There are many (possibly an infinite number) of samples of size n. With our one sample, we estimate one set of parameter values. For each sample, we get a different set of estimated parameter values (we get different parameter values for each sample using the same distribution).

5.5 Weights

For linear parameter estimation it has been shown that optimum weights for the observations are given by the inverse of the covariance matrix of the uncertainties. The uncertainties cannot always be found since this involves knowing the true values of the data (measurement error = zero) as well as the uncertainties due to sample size. Instead, we will use the inverse of the covariance matrix of the residuals. For data which has small measurement error, the results using the two methods should be close.

[39] Equations (5-9) are called the "normal equations".

For our nonlinear parameter estimation we will adopt the same approach even though it has not been proven to be optimal: we assume that optimal weights are given by the inverse covariance matrix of the residuals.

The major contributor to the residuals is assumed to be the effect of variability due to sample size ($\varepsilon_{\mu 3}$). For this reason the most general weights we will use are the inverse of the covariance matrix of the order statistics [eq (4-7c)].

5.6 Assignment of Probability

We need a method to transform our order numbers $r = 1, 2, \ldots, n$ into probability.

In the development of our models in section 5.3 we assumed that we had the values for both independent and dependent variables. For example, we might obtain a sample of bending moment values from a towing tank test. At first glance, there are no corresponding variables to relate them to. What then do we relate them to? The answer is that each of them corresponds to a particular probability in a specified distribution. If we knew the distribution type and distribution parameters, we could compute the probability for each value. However, the distribution parameter values are what we are trying to estimate - they are unknown. In addition, sometimes we are trying to sort out which distribution to use.

There has been much effort expended on the assignment of probability (this is known as the "plotting position problem"). Kimball [1960] and Gumbel [1958, §1.2.6, 1.2.7] are two references. ***In this report we will use $P_r = r/(n+1)$.*** This is a distribution free method (using medians, modes, etc. requires that we know the distribution as well as the distribution parameter values - a distribution specific iterative process which is computationally intensive). This choice of the plotting position can be shown to represent the average value of probability for the order statistics of a uniform distribution. This result is important since any distribution may be transformed into a uniform distribution by means of the probability integral transformation: "... this result implies that the order statistics divide the area under the curve $y = p(x)$ into $n+1$ parts, each with expected value $1/(n+1)$." [David, 1970, §3.1].

Note, however, that we have added yet another source of uncertainty - each plotting position choice for P has its own sampling distributions for estimated parameter values.

5.7 Objective Function Example

The Weibull distribution is given by

$$P = 1 - e^{-\left\{\left[(x-x_0)/sf\right]^c\right\}} \qquad (1\text{-}1)$$

When $x_0 = 0$ this may be solved for x (Appendix B) to obtain

$$x = sf * \left[\ln\left(1/(1-P)\right)\right]^{1/c} \qquad (\text{B-10})$$

Suppose we have three observed data points (measured with negligible uncertainty) d_1, d_2, and d_3 which have been sorted in size from smallest to largest. The corresponding $P_r = r/(n+1)$ are 1/(3+1), 2/(3+1), and 3/(3+1) = 0.25, 0.50, and 0.75.

The residual vector is

$$\mathbf{e}_{\boldsymbol{\mu}}(\boldsymbol{\theta})=\begin{bmatrix} d_1 - x_1 \\ d_2 - x_2 \\ d_3 - x_3 \end{bmatrix} = \begin{bmatrix} d_1 - \left(sf * \left[\ln(1/(1-P_1))\right]^{1/c} \right) \\ d_2 - \left(sf * \left[\ln(1/(1-P_2))\right]^{1/c} \right) \\ d_3 - \left(sf * \left[\ln(1/(1-P_3))\right]^{1/c} \right) \end{bmatrix} = \begin{bmatrix} d_1 - \left(sf * \left[\ln(1/(1-0.25))\right]^{1/c} \right) \\ d_2 - \left(sf * \left[\ln(1/(1-0.50))\right]^{1/c} \right) \\ d_3 - \left(sf * \left[\ln(1/(1-0.75))\right]^{1/c} \right) \end{bmatrix} \quad (5\text{-}10a)$$

where $$\boldsymbol{\theta} = \begin{bmatrix} sf \\ c \end{bmatrix} \quad (5\text{-}10b)$$

If all weights are = 1 in eq (5-8b) then V = identity matrix so we have

$$\begin{aligned} \text{objf} &= \mathbf{e}_{\boldsymbol{\mu}}(\boldsymbol{\theta})^{\mathbf{T}} * \mathbf{e}_{\boldsymbol{\mu}}(\boldsymbol{\theta}) = \left[(d_1 - x_1) \quad (d_2 - x) \quad (d_3 - x_3) \right] * \begin{bmatrix} d_1 - x_1 \\ d_2 - x_2 \\ d_3 - x_3 \end{bmatrix} \\ &= (d_1 - x_1)^2 + (d_2 - x_2)^2 + (d_3 - x_3)^2 \end{aligned} \quad (5\text{-}11)$$

which is the (scalar) sum of squares having unknown values for "sf" and "c".

If we weight only by the inverse of the variances for r = 1, 2, 3 we have

$$\begin{aligned} \text{objf} &= \mathbf{e}_{\boldsymbol{\mu}}(\boldsymbol{\theta})^{\mathbf{T}} * \mathbf{VI} * \mathbf{e}_{\boldsymbol{\mu}}(\boldsymbol{\theta}) \\ &= \left[(d_1 - x_1) \quad (d_2 - x) \quad (d_3 - x_3) \right] * \begin{bmatrix} VI_{1,1} & 0 & 0 \\ 0 & VI_{2,2} & 0 \\ 0 & 0 & VI_{3,3} \end{bmatrix} \begin{bmatrix} d_1 - x_1 \\ d_2 - x_2 \\ d_3 - x_3 \end{bmatrix} \\ &= VI_{1,1} * (d_1 - x_1)^2 + VI_{2,2} * (d_2 - x_2)^2 + VI_{3,3} * (d_3 - x_3)^2 \end{aligned} \quad (5\text{-}12)$$

Here $VI_{r,r} = 1/V_{r,r}$

where $V_{1,1}$ variance of x for $r = 1$
$V_{2,2}$ variance of x for $r = 2$
$V_{3,3}$ variance of x for $r = 3$

The entries of the variance matrix inverse when the covariances = 0 are given by 1/variance.

The variances are given by eqs (4-7a) which in turn need the mean and the probability density of the r^{th} order statistic in a sample of size "n". The means are given by eqs (4-8a). The probability density of the r^{th} in a sample of size "n" is given by eq (4-1).

Note that the values of the variances depend upon the parameters (sf, c) of the distribution. This requires an iterative process since, at each step of the (nonlinear) solution for (sf, c), new variance values must be computed which are consistent with the solution values for (sf, c).

Weighting using the full inverse covariance matrix of the order statistics results in

$$\begin{aligned} \text{objf} &= \mathbf{e}_{\boldsymbol{\mu}}(\boldsymbol{\theta})^{\mathbf{T}} * \mathbf{V}^{-1} * \mathbf{e}_{\boldsymbol{\mu}}(\boldsymbol{\theta}) \\ &= \left[(d_1 - x_1) \quad (d_2 - x) \quad (d_3 - x_3)\right] * \begin{bmatrix} V_{1,1} & V_{1,2} & V_{1,3} \\ V_{2,1} & V_{2,2} & V_{2,3} \\ V_{3,1} & V_{3,2} & V_{3,3} \end{bmatrix}^{-1} \begin{bmatrix} d_1 - x_1 \\ d_2 - x_2 \\ d_3 - x_3 \end{bmatrix} \end{aligned} \tag{5-13}$$

Here **V** is the covariance matrix of the order statistics. The covariance matrix is symmetric so that $V_{i,j} = V_{j,i}$ thus reducing the computational work. After its entries are found its inverse must be computed.

The variances are given by eqs (4-7a) which in turn need the mean and the probability density of the r^{th} in a sample of size "n". The covariances are given by eqs (4-7b) which in turn need the mean and the probability density of the r^{th} and s^{th} in a sample of size "n". The means are given by eqs (4-8a). The probability density of the r^{th} and s^{th} in a sample of size "n" is given by eq (4-9).

Again, note that the values of the variances depend upon the parameters (sf, c) of the distribution. This requires an iterative process since, at each step of the (nonlinear) solution for (sf, c), new variance values must be computed which are consistent with the solution values for (sf, c).

It seems plausible that the integrations for the variances and covariances could be performed on a spreadsheet. If the spreadsheet can also do a matrix inverse along with having a nonlinear equation solver then the entire parameter estimation can be done on a spreadsheet.[40]

5.8 Weibull Distribution Examples

The use of the Weibull distribution to illustrate some of the ideas presented here should not be taken as an endorsement that the Weibull distribution is the distribution of choice for data. Other distributions may do a better job of fitting the data, and, if so, should be used. The Weibull distribution is particularly analytically tractable and suitable for these examples.

---

[40] For the examples shown here a program capable of doing symbolic mathematics was used in conjunction with a spreadsheet.

Recall that the Weibull distribution is given by

$$P = 1 - e^{-\left\{\left[(x-x_0)/sf\right]^c\right\}} \tag{1-1}$$

where

P	probability that the value of the variable is x or less
x	variable such as load or response
sf	scale factor (characteristic value = sf + x_0)
x_0	truncation value (value below which the probability is always zero)
c	exponent (also called the slope)

In original space we use the data as collected: no data transformations are performed. Even for methods which do not weight the data, sorting is necessary so that we may assign a probability to each observation. For methods which use weights, or which use transforms to some other space, such as Weibull, Gaussian, lognormal, etc., ordering of the data by size, usually from smallest to largest, is necessary. Note that whenever the data are sorted by size, application of order statistics is possible.

We will estimate the parameters using five different methods for a data set of nine points:

a) fit in Weibull space, use ln(data) as the independent variable;

b) fit in Weibull space, use lnln(1/(1-probability)) as the independent variable;

c) fit in original space, use probability as the independent variable, all weights = 1;

d) fit in original space, use probability as the independent variable, weights = 1/variance found from order statistics;

e) fit in original space, use probability as the independent variable, weights = inverse of the covariance matrix found from order statistics.

The above five different methods are ranked from the least desirable "a" (section 5.8.2) to the most desirable "e" (section 5.8.6). Comments on the desirability of a method are in the section showing the method. Recall that our criterion, in addition to estimating parameter values, is the ability to use our fitting method to aid in estimating the parameter covariance matrix.

Method "a", despite demonstrations such as in Richardson [1992] that other methods are superior, is still very widely used because, in part: a) linearity in Weibull space can be easily checked visually, b) the Weibull slope estimate can be directly read from the Weibull plot, and c) the computations for slope are easy in the sense that they are linear.

As the sample size becomes larger, the differences in estimated parameter values between the different methods become smaller.

The methods illustrated here all deal directly with the data. Methods such as estimating the Weibull parameters using moments of the data, such as the method of moments, and its associated parameter covariance matrix estimate, are not discussed here (but should be a topic of future work).

A principal reason for using methods which deal directly with the data is that using a weight for each data point is straightforward. This leads to the possibility of making more accurate estimates of the parameter covariance matrix than can be obtained by other methods. This advantage becomes more important as the sample size becomes smaller.

The parameter covariance matrix provides an indication of how tight are our parameter estimates. This applies whether the parameter covariance matrix is estimated using the method shown in section 5.9 or by simulation.

The parameter covariance matrix is one of the two pieces of information needed to estimate the uncertainty of a value computed from the fit (the other is the sensitivity of the fit to changes in parameter values - see section 9.5).

The data for the examples are given in Table 5-1.

Table 5-1. Data for Examples

r	value
1	17654
2	20975
3	21808
4	24687
5	24701
6	28872
7	34064
8	39526
9	40706

<u>Comments</u>

1) This is a small data set, and so will have correspondingly large parameter covariance estimates. To obtain smaller parameter covariance estimates we need to obtain a larger data set.

2) Small data sets arise when the phenomena producing the data set occur infrequently. If the data values in the set have large magnitudes, they cannot be ignored. The difficulties in selecting a distribution and estimating parameter values are particularly severe in these situations.

3) As an example, when slamming is present the probability of occurrence of a small data set is high. If the probability of slamming is 10% (a very high slamming probability), then a sample of 180 events will, on the average, contain 18 slams. These 18 slams will often be the 18 largest events, and usually plot on Weibull paper with a different slope than the remaining 162 = (180 - 18) events: i.e., they constitute a different distribution than the distribution which describes the remaining 162 points. If the probability of slamming is 5% (a more reasonable probability), then a sample of 180 events will, on the average, contain 9 slams. These 9 slams will often be the 9 largest events, and usually plot on Weibull paper as a distribution with a

different slope than the remaining 171 = (180 - 9) events. Even though we have collected a total sample of reasonable size (180), the subsample of interest, in this case of slams, may be small. Note that in order to double the number of occurrences of infrequent events, in this case slams, we need to double the total test time - often not possible due to schedule and budget constraints.

4) Small data sets exhibit variability in their size from run to run. Nine slams in a data set of 180 is the expected number of events in the small data set when p = 0.05. This is the average number of events in the small data set when many sets of the total data set (180) are obtained. The probability of getting exactly "m" events in the small data set is given by the binomial distribution term $C_m^n * p^m * (1-p)^{n-m}$ where $C_m^n = n!/[m!\ (n-m)!]$. For the case where we expect 5% of 180 events to be slams, the distribution of the number of slams is given by $180!/[m!\ (180-m)!] * 0.05^m * (1-0.05)^{180-m}$ for m = 0, 1, 2, ... , 180. A plot of the probability mass function of a sample of size of "m" in a total sample of 180, for P average = 0.01, 0.02, 0.05, and 0.10 is shown in Figure 5-1.

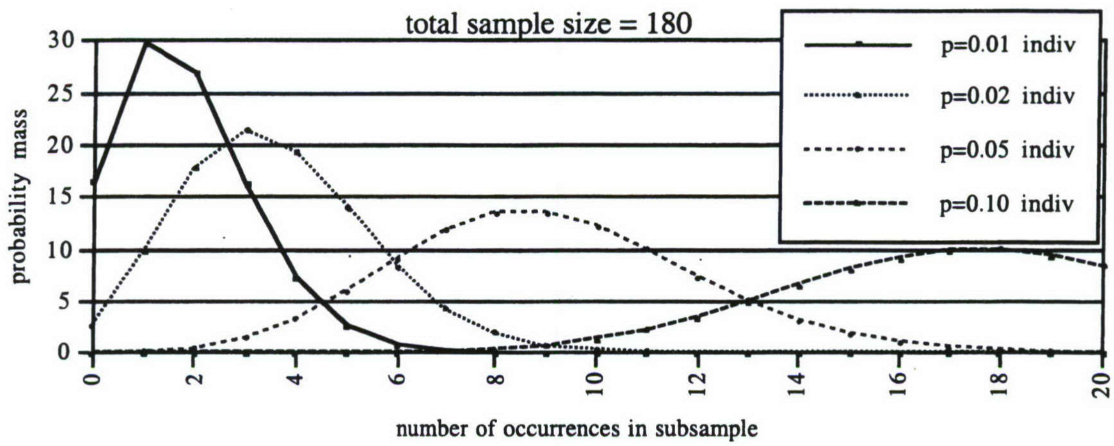


Figure 5-1. Probability of Occurrence of Subsample Size "m"

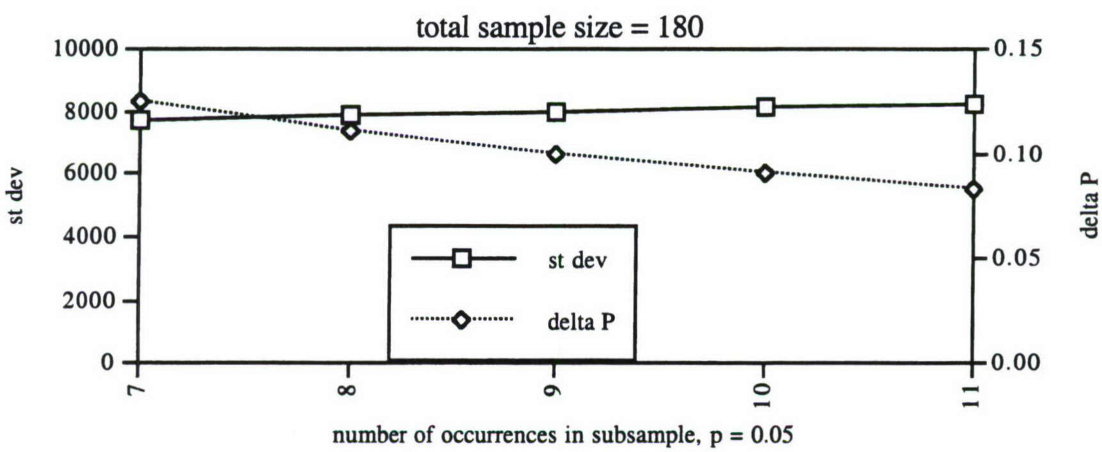


Figure 5-2. Standard Deviation and Change in P vs. Subsample Size "m"

Figure 5-2 shows the standard deviation (left hand axis) and the change in plotting probability increment (right hand axis) as a function of subsample size. For example, the change in probability increment for subsample size 7 is given by 2/(7+1) - 1/(7+1) = 2/8 - 1/8 = 1/8 = 0.125. [A Weibull distribution with scale factor = 31,400 and slope 3.25 was used to obtain the standard deviation values. These scale factor and slope values are the averages of the values shown in Table 5-7 (excluding Dubey) for the five methods "a" through "e"].

That the standard deviations are increasing with subsample size, while the probability increments are decreasing, shows that taking probability as the independent variable will lead to decreasing variability with increasing sample size. Taking data values as independent results in increasing variability with increased sample size. Hence the conclusion that method "a" is the least desirable method for estimating Weibull parameter values.

The change in standard deviation of 6% from a sample size of 7 to one of 11 suggests that the standard deviation is effectively stable (the change in average was negligible) for estimating purposes. The major variability in small subsample sizes is due to sampling variability for a particular subsample size such as 7.

5.8.1 Getting Started

The three nonlinear estimation methods (section 5.8 methods c, d, and e) require initial parameter values to start the procedure to find the solution values. The two methods (section 5.8 methods "a" and "b" which use transforms (mappings) of the data into Weibull space do not require initial values since they are linear in the transformed parameters.

Assuming the data follows a Weibull distribution, two methods not requiring iteration to obtain initial estimates are:

a) Use the method of Dubey (estimates all three Weibull distribution parameters).

b) Estimate the slope and scale factor using a plot in Weibull space. However, to make a plot in Weibull space requires that we should somehow estimate x_0 .

Data are transformed (mapped) into Weibull space by computing $\ln(x-x_0)$ and $\ln(\ln(1/(1-P_r)))$ [Appendix B] where $P_r = r/(n+1)$. Under this transform the transformed data will plot as a straight line if they follow a Weibull distribution. A Weibull plot of the data using $x_0 = 0$ is shown in Figure 5-3.

For the Weibull plot of Figure 5-3, the transformed data are fitted by the straight line y = m*x + b = 3.273*x - 33.896. The slope "c" is given by m = 3.273. The scale factor is given (Appendix B) by sf = e**(-b/m) = e**(-(-33.896)/3.273) = 31453.

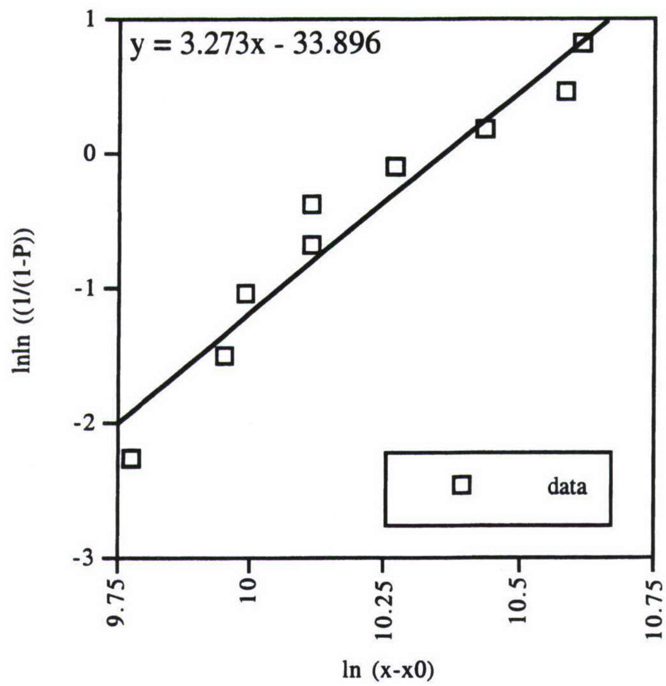


Figure 5-3. Weibull Space Plot of Data for Fitting Examples

5.8.1.1 Method of Dubey

Formulas due to Dubey for initial estimates of Weibull parameters are given in Castillo [1988]:[41]

$$x_0 = \frac{X_{(1)}X_{(k)} - X_{(j)}^2}{X_{(1)} + X_{(k)} - 2X_{(j)}} \tag{5-16a}$$

$$c = \frac{2.99}{\ln\left(X_{[0.9737n+1]} - x_0\right) - \ln\left(X_{[0.1673n+1]} - x_0\right)} \tag{5-16b}$$

$$sf = \frac{X_{[0.1673n+1]} - x_0}{0.183^{1/c}} \tag{5-16c}$$

where

$X_{(1)} < X_{(j)} < \left(X_{(1)}X_{(k)}\right)^{1/2}$

and [ ] indicates integer part

The nine data points were shown in Table 5-1.

41 Notation and equation numbers changed to be consistent with this report.

Using k=8, we have $\left(X_{(1)}X_{(k)}\right)^{1/2} = (17654 * 39526)^{1/2} = 26416$. 24687 (Table 5-1, j = 4), while slightly greater than $\left(X_{(1)}X_{(k)}\right)^{1/2}$, will be used.

$$x_0 = \frac{X_{(1)}X_{(k)} - X_{(j)}^2}{X_{(1)} + X_{(k)} - 2X_{(j)}} = \frac{17654 * 39526 - 24687^2}{17654 + 39526 - 2 * 24687} = 11317$$

$$[0.9737n + 1] = [0.9737 * 9 + 1] = [9.7633] = 9$$

$$[0.1673n + 1] = [0.1673 * 9 + 1] = [2.5057] = 2$$

$$c = \frac{2.99}{\ln\left(X_{(9)} - x_0\right) - \ln\left(X_{(2)} - x_0\right)} = \frac{2.99}{\ln(40706 - 11317) - \ln(20975 - 11317)} = 2.687$$

$$sf = \frac{X_{[0.1673n+1]} - x_0}{0.183^{1/c}} = \frac{X_{(2)} - x_0}{0.183^{1/c}} = \frac{20975 - 11317}{0.183^{1/2.687}} = 18171$$

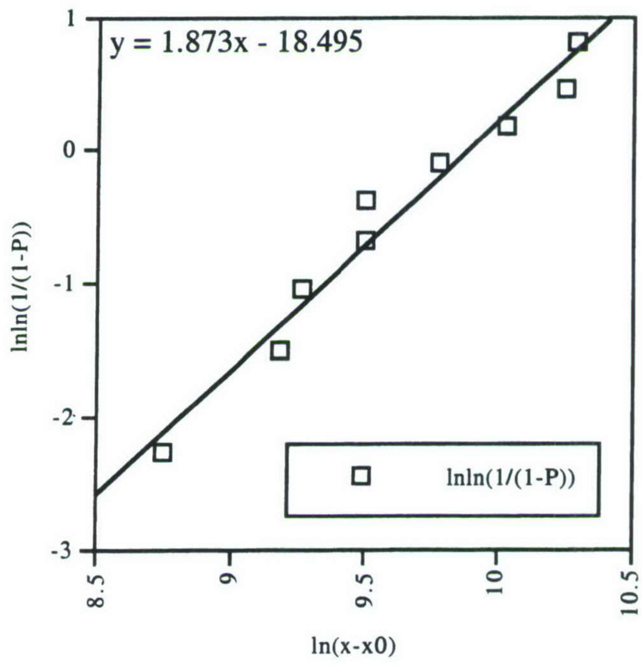


Figure 5-4. Weibull Plot Using Dubey Estimate for x_0

The Dubey estimated slope = 2.687 in original space results in an estimated slope = 1.873 in Weibull space. Note that the possible curvature apparent in Figure 5-3 has largely been removed in Figure 5-4 due to x_0 not being equal to 0.

Well, why don't we use the Dubey estimates as starting values? The reason is that we know from physical considerations that the value for x_0 is zero (assuming a zero mean process) since the data represent bending moments. We repeat the Dubey calculations using $x_0 = 0$ to obtain

$$c = \frac{2.99}{\ln\left(X_{(9)} - x_0\right) - \ln\left(X_{(2)} - x_0\right)} = \frac{2.99}{\ln(40706 - 0) - \ln(20975 - 0)} = 4.510$$

$$sf = \frac{X_{[0.1673n+1]} - x_0}{0.183^{1/c}} = \frac{X_{(2)} - x_0}{0.183^{1/c}} = \frac{20975 - 0}{0.183^{1/4.510}} = 30566$$

The next five pages show estimation of the parameters using five different methods for the data set of nine points:

a) fit in Weibull space, use ln(data) as the independent variable;

b) fit in Weibull space, use lnln(1/(1-probability)) as the independent variable;

c) fit in original space, use probability as the independent variable, all weights = 1;

d) fit in original space, use probability as the independent variable, weights = 1/variance found from order statistics;

e) fit in original space, use probability as the independent variable, weights = inverse of the covariance matrix found from order statistics.

The above five different methods are ranked from the least desirable "a" (section 5.8.2) to the most desirable "e" (section 5.8.6). Comments on the desirability of a method are in the section showing the method. Recall that our criterion, in addition to estimating parameter values, is the ability to use our fitting method to aid in estimating the parameter covariance matrix.

5.8.2 ln(data-x_0) Assumed Independent

Since the data of Table 5-2 consist of bending moments x_0 was set to zero since it is reasonable that the smallest bending moment value can be zero. "indepen" indicates which variable is taken as independent, while "depend" indicates the dependent data and the fit.

The top part of Table 5-2 shows the solution for fitting in Weibull space assuming ln(x-x_0) [col 3] is independent. Scale factor "sf" and slope "c" were found using linear least squares in Weibull space. The bottom part of Table 5-2 shows the data (col 1), its transforms (cols 2, 3, and 6), and the fit (col 8). The fit (col 8) are the ln(ln(1/(1-P))) values corresponding to the ln (data-x_0) values (col 3).

The dependent variable is "ln(ln(1/(1-P)))" [col 6]. The residual [col 9] is "data ln(ln(1/(1-P))) - fit ln(ln(1/(1-P))))" = [(col 6) - (col 8)]. The objective function = "objf func" = 5.00E-01 = sum{ residual^2 [col 10] }.

"sd av res" is the standard deviation of the average residual. "sd av err" is an approximation to the standard deviation of the average error. The "av res" has n-1 degrees of freedom, while "av err" has n-2 degrees of freedom since 2 parameters were estimated.

Table 5-2. Fit in Weibull Space, ln sag Independent

		x_0	sf	c	objf func	sd av res	sd av err	sf, c vary	
		0	31437	3.2732	5.00E-01	0.250	0.267	2	
col 1 sag data sorted	col 2 sag -x_0	col 3 ln (data -x_0) indepen	col 4 r	col 5 P data = r/10	col 6 lnln 1/ (1-P) depend	col 7 P fit	col 8 lnln P fit depend	col 9 resid in lnln P	col 10 (resid in lnln P)^2
17654	17654	9.7787	1	0.1000	-2.2504	0.1404	-1.8887	-0.3617	1.31E-01
20975	20975	9.9511	2	0.2000	-1.4999	0.2335	-1.3245	-0.1754	3.08E-02
21808	21808	9.9900	3	0.3000	-1.0309	0.2607	-1.1970	0.1661	2.76E-02
24687	24687	10.1140	4	0.4000	-0.6717	0.3645	-0.7912	0.1194	1.43E-02
24701	24701	10.1146	5	0.5000	-0.3665	0.3650	-0.7893	0.4228	1.79E-01
28872	28872	10.2706	6	0.6000	-0.0874	0.5309	-0.2786	0.1912	3.65E-02
34064	34064	10.4360	7	0.7000	0.1856	0.7276	0.2627	-0.0771	5.94E-03
39526	39526	10.5847	8	0.8000	0.4759	0.8795	0.7495	-0.2736	7.48E-02
40706	40706	10.6141	9	0.9000	0.8340	0.9027	0.8458	-0.0117	1.37E-04
								sum =	5.00E-01

This is the least desirable way to estimate Weibull parameters for two reasons:

a) when using linear least squares (as we are doing) taking x-x_0 as independent means we assume that there is no error in the bending moment values. This is equivalent to assuming that another sample of 9 points will have the same values - an assumption contradicted both by experience and by the computed variability found by using order statistics;

b) using ln(x-x_0) means that the fit is driven by the smallest values since the log of a small number has more variability than the log of a large number.

5.8.3 ln(ln(1/(1-P_m))) Assumed Independent

Since the data of Table 5-3 consist of bending moments x_0 was set to zero since it is reasonable that the smallest bending moment value can be zero.

The top part of Table 5-3 shows the solution for fitting in Weibull space assuming ln(ln(1/(1-P))) [col 6] is independent. Scale factor "sf" and slope "c" were found using linear least squares in Weibull space. The bottom part of Table 5-3 shows the data (col 1), its transforms (cols 2, 3, and 6), and the fit (col 8). The fit (col 8) are the ln(sag-x_0) values corresponding to the ln(ln(1/(1-P))) values (col 6).

The dependent variable is "ln(data-x_0)" [col 3]. The residual [col 9] is "ln(data-x_0) - ln(fit-x_0)" [(col 3) - (col 8)]. The objective function = "objf func" = 4.36E-02 = sum{ residual^2 = [col 10] }.

Table 5-3. Fit in Weibull Space, lnln(1/(1-P)) Independent

x_0	sf	c	objf func	sd av res	sd av err	sf, c vary
0	31136	3.4983	4.36E-02	0.074	0.079	2

col 1 sag data sorted	col 2 sag -x_0	col 3 ln (data -x_0) depend	col 4 r	col 5 P data = r/10	col 6 lnln P data indepen	col 7 sag (fit -x_0)	col 8 ln (sag -x_0) fit depend	col 9 diff in ln x	col 10 (diff in ln x)^2
17654	17654	9.7787	1	0.1000	-2.2504	16364	9.7028	0.0759	5.76E-03
20975	20975	9.9511	2	0.2000	-1.4999	20279	9.9173	0.0337	1.14E-03
21808	21808	9.9900	3	0.3000	-1.0309	23188	10.0514	-0.0614	3.77E-03
24687	24687	10.1140	4	0.4000	-0.6717	25696	10.1541	-0.0401	1.60E-03
24701	24701	10.1146	5	0.5000	-0.3665	28039	10.2413	-0.1267	1.61E-02
28872	28872	10.2706	6	0.6000	-0.0874	30367	10.3211	-0.0505	2.55E-03
34064	34064	10.4360	7	0.7000	0.1856	32832	10.3992	0.0368	1.36E-03
39526	39526	10.5847	8	0.8000	0.4759	35673	10.4821	0.1026	1.05E-02
40706	40706	10.6141	9	0.9000	0.8340	39518	10.5845	0.0296	8.77E-04
									sum = 4.36E-02

This is the second least desirable way to estimate Weibull parameters for two reasons:

a) when using linear least squares (as we are doing) taking P as independent means we assume that there is no error in the sample size. This is equivalent to assuming that another set of runs down the tank will again produce a sample of nine points - an assumption contradicted by experience. However, as sample size increases, the effect of the variability due to the number of points will decrease, whereas the variability in the data magnitudes will not change very much [Table 4-2].

b) using ln(x-x_0) means that the fit is driven by the smallest values since the log of a small number has more variability than the log of a large number.

5.8.4 P Independent, Wts = 1

Since the data of Table 5-4 consist of bending moments x_0 was set to zero since it is reasonable that the smallest bending moment value can be zero.

The top part of Table 5-4 shows the solution for fitting in original space using weights = 1 assuming P [col 5] is independent. Scale factor "sf" and slope "c" were found using a nonlinear equation solver in original space.

The bottom part of Table 5-4 shows the data, (col 1), and the fit (col 7). The fit (col 7) are the sag values corresponding to the probability values (col 5). The dependent variable is "sag - x_0" [col 2]. The residual [col 9] is "(data-x_0) - (fit-x_0)" [(col 2) - (col 7)]. The objective function = "objf func" = 1.58E+00 = sum[residual^2 = (col 9)^2].

Cols 3, 6, and 8 do not enter into the solution.

Table 5-4. Fit in Original Space, P Independent, All Wts = 1

x_0	sf	c	objf func	sd av resid	sd av err	sf, c vary
0	31450	3.1578	3.15E+07	1985	2122	2

col 1 sag data sorted	col 2 sag data - x_0 depend	col 4 r	col 5 P data = r/10 indepen	col 7 sag fit depend	col 9 diff = data - fit	col 10 diff^2
17654	17654	1	0.1000	15421	2233	4.99E+06
20975	20975	2	0.2000	19558	1417	2.01E+06
21808	21808	3	0.3000	22690	-882	7.78E+05
24687	24687	4	0.4000	25423	-736	5.42E+05
24701	24701	5	0.5000	28003	-3302	1.09E+07
28872	28872	6	0.6000	30591	-1719	2.96E+06
34064	34064	7	0.7000	33354	710	5.04E+05
39526	39526	8	0.8000	36565	2961	8.77E+06
40706	40706	9	0.9000	40957	-251	6.29E+04
					sum =	3.15E+07

This is a relatively quick way to estimate Weibull parameters, since, by staying in original space, we have no problems associated with transforming (mapping) into some other space.

However, taking P as independent means we are assuming that there is no error in the sample size, i.e., that another run down the tank will again produce a sample of nine points - an assumption contradicted by experience. However, as sample size increases, the effect of the variability due to the number of points will decrease, whereas the variability in the data magnitudes will not change very much [Table 4-2].

We assume here that all data weights = 1. This assumption is contradicted by the results from order statistics which show different weights for different order numbers.

5.8.5 *P Independent, Wts = Var* $^{-1}$

Since the data of Table 5-5 consist of bending moments x_0 was set to zero since it is reasonable that the smallest bending moment value can be zero.

The top part of Table 5-5 shows the solution for fitting in original space using weights = 1/variance assuming P [col 5] is independent. Scale factor "sf" and slope "c" were found using a nonlinear equation solver in original space. The variances were recomputed at each nonlinear solution until the solution values of (sf, c) were consistent with the (sf, c) values used to compute the variances.

The bottom part of Table 5-5 shows the data, (col 1), and the fit (col 7). The fit (col 7) are the sag values corresponding to the probability values (col 5). The dependent variable is "sag - x_0" [col 2]. The residual [col 9] is "(data-x_0) - (fit-x_0)" [(col 2) - (col 7)]. The objective function = "objf func" = 1.58E+00 = sum{ wt*(residual^2) = (col 11)*[(col 9)^2] }.

Cols 3, 6, and 8 do not enter into the solution.

Table 5-5. Fit in Original Space, P Independent, Wts = variance^{-1}

	x_0	sf	c	objf func	sd av res	sd av err	sf, c vary	consistent
	0	31292	3.1206	1.58E+00	1993	2130	2	yes

col 1 sag data sorted	col 2 sag - x_0 depend	col 4 r	col 5 P data = r/10 indepen	col 7 sag fit depend	col 9 diff data - fit	col 10 var	col 11 1/var (wt)	col 12 (diff^2) * wt
17654	17654	1	0.1000	15214	2440	2.36E+07	4.24E-08	2.52E-01
20975	20975	2	0.2000	19350	1625	1.98E+07	5.05E-08	1.33E-01
21808	21808	3	0.3000	22488	-680	1.81E+07	5.53E-08	2.56E-02
24687	24687	4	0.4000	25232	-545	1.73E+07	5.77E-08	1.71E-02
24701	24701	5	0.5000	27824	-3123	1.72E+07	5.81E-08	5.66E-01
28872	28872	6	0.6000	30428	-1556	1.78E+07	5.62E-08	1.36E-01
34064	34064	7	0.7000	33210	854	1.93E+07	5.18E-08	3.78E-02
39526	39526	8	0.8000	36447	3079	2.29E+07	4.37E-08	4.14E-01
40706	40706	9	0.9000	40880	-174	3.44E+07	2.91E-08	8.76E-04
							sum =	1.58E+00

This is a good way to estimate Weibull parameters, since, by staying in original space, we a) have no problems associated with transforming (mapping) into some other space, and b) allow for the possibility of estimating parameter covariances.

However, taking P as independent means we are assuming that there is no error in the sample size, i.e., that another run down the tank will again produce a sample of nine points - an assumption contradicted by experience. However, as sample size increases, the effect of the variability due to the number of points will decrease, whereas the variability in the data magnitudes will not change very much [Table 4-2].

5.8.6 P Independent, Wts = (covariance matrix)$^{-1}$

Since the data of Table 5-6 consist of bending moments x_0 was set to zero since it is reasonable that the smallest bending moment value can be zero.

Table 5-6. Fit in Original Space, P Independent, Wts = (covar matrix)$^{-1}$

x_0	sf	c	objf func	sd av resid	sd av err	sf, c vary
0	31585	3.2084	3.75E+00	1993	2130	2

col 1 sag data sorted	col 2 sag - x_0 depend	col 4 r	col 5 P data = r/10 indepen	col 7 sag fit depend	col 9 diff data - fit
17654	17654	1	0.1000	15663	1991
20975	20975	2	0.2000	19790	1185
21808	21808	3	0.3000	22905	-1097
24687	24687	4	0.4000	25619	-932
24701	24701	5	0.5000	28175	-3474
28872	28872	6	0.6000	30736	-1864
34064	34064	7	0.7000	33466	598
39526	39526	8	0.8000	36635	2891
40706	40706	9	0.9000	40962	-256

Table 5-6 (cont.)

col 10 1	col 11 2	col 12 3	col 13 4	col 14 5	col 15 6	col 16 7	col 17 8	col 18 9
1991	1185	-1097	-932	-3474	-1864	598	2891	-256
7.20E-08	-5.14E-08	1.92E-10	1.02E-10	5.84E-11	3.50E-11	2.10E-11	1.20E-11	5.42E-12
-5.14E-08	1.55E-07	-9.48E-08	2.16E-10	1.28E-10	7.83E-11	4.80E-11	2.80E-11	1.29E-11
1.92E-10	-9.48E-08	2.31E-07	-1.28E-07	2.19E-10	1.38E-10	8.67E-11	5.17E-11	2.43E-11
1.02E-10	2.16E-10	-1.28E-07	2.80E-07	-1.45E-07	2.27E-10	1.47E-10	9.00E-11	4.36E-11
5.84E-11	1.28E-10	2.19E-10	-1.45E-07	2.96E-07	-1.43E-07	2.46E-10	1.56E-10	7.84E-11
3.50E-11	7.83E-11	1.38E-10	2.27E-10	-1.43E-07	2.75E-07	-1.23E-07	2.81E-10	1.48E-10
2.10E-11	4.80E-11	8.67E-11	1.47E-10	2.46E-10	-1.23E-07	2.19E-07	-8.69E-08	3.13E-10
1.20E-11	2.80E-11	5.17E-11	9.00E-11	1.56E-10	2.81E-10	-8.69E-08	1.36E-07	-3.96E-08
5.42E-12	1.29E-11	2.43E-11	4.36E-11	7.84E-11	1.48E-10	3.13E-10	-3.96E-08	5.09E-08

The top part of Table 5-6 shows the solution for fitting in original space using weights = (covariance matrix)$^{-1}$ assuming P [col 5] is independent. Scale factor "sf" and slope "c" were found using a nonlinear equation solver in original space. The covariance matrix and its inverse were recomputed at each nonlinear solution until the solution values of (sf, c) were consistent with the (sf, c) values used to compute the covariances.

The bottom part of Table 5-6 shows the data, (col 1), and the fit (col 7). The fit (col 7) are the sag values corresponding to the probability values (col 5). The dependent variable is "sag - x_0" [col 2]. The residual [col 9] is "(data-x_0) - (fit-x_0)" [(col 2) - (col 7)]. The objective function = "objf func" = 3.75E+00 = sum{ residual^T * VI *residual. The residual is in col 9 enclosed in double lines. Its transpose (residual^T) is in the top row of Table 5-5 (cont.) enclosed in double lines. The inverse of the covariance matrix occupies the bottom of Table 5-5 (cont.).

Cols 3, 6, and 8 do not enter into the solution.

This is the best way to estimate Weibull parameters, since, by staying in original space, we a) have no problems associated with transforming (mapping) into some other space, and b) allow for the possibility of being able to estimate parameter covariances.

However, taking P as independent means we are assuming that there is no error in the sample size, i.e., that another run down the tank will again produce a sample of nine points - an assumption contradicted by experience. However, as sample size increases, the effect of the variability due to the number of points will decrease, whereas the variability in the data magnitudes will not change very much [Table 4-2].

5.8.7 Weibull Plot Parameter Bias

Very preliminary results from work currently being performed under another task suggests that there is bias in the estimated slope ("c") values when Weibull plots are used to estimate slopes. This bias appears to be more severe for smaller samples than for larger samples. Whether the bias is also a function of slope is also being investigated. There does not appear to be much bias in the scale factor "sf" estimate when Weibull plots are used. If slope bias is confirmed, this becomes yet another reason to estimate parameter values in original, not Weibull space.

5.8.8 Fitting Examples Summary

For this particular example the estimates of "sf" and "c" do not vary much (Table 5-7). This is quite unusual for so small a sample size (n = 9).

Table 5-7. Summary of Weibull Parameter Estimation Examples

	DuBey (x_0 = 0)	method a (Table 5-2)	method b (Table 5-3)	method c (Table 5-4)	method d (Table 5-5)	method e (Table 5-6)
independ ->		ln(x-x_0)	lnln(1/(1-P))	P	P	P
can param cov be est?	no	yes	yes	yes	yes	yes
sf	30600	31400	31100	31500	31300	31600
c	4.510	3.27	3.50	3.16	3.12	3.21

Parameter covariance matrices for linear methods "a" and "b" would be computed using the Weibull transformed values "ln(data-x_0)" and "ln(ln(1/(1-P)))". The results would be in terms of slope or 1/slope and intercept, not scale factor. A minimum variance bound formula for the parameter covariance matrix is known when linear parameter estimation is done.

Any parameter covariance estimates computed in Weibull space would have to be transformed into original space since this is the space in which lifetime load estimates, and lifetime load uncertainty estimates are made.

Parameter covariance matrices for nonlinear methods "c", "d" and "e" (sections 5.8.4, 5.8.5, and 5.8.6) would be computed using the data in original space. A minimum variance bound formula for the parameter covariance matrix is not known when nonlinear parameter estimation is done. Parameter covariance matrices for methods "d" and "e" (sections 5.8.5 and 5.8.6) only will be shown in the next section. (Methods "a", "b", and "c" (sections 5.8.2, 5.8.3, and 5.8.4) are not shown).

5.9 Parameter Covariance Estimate: Nonlinear Case

We not only want to estimate the parameters, but would like to have some idea of how "good" are these values. The best (and very, very time consuming) way to estimate the distribution of the estimated parameter values is to do many simulations and make histograms of the resulting parameter values. From these histograms we may find, for example, the 1% and 99% tolerance limits[42] (these would cover 98% of all parameter values). In addition, we can estimate the parameter variances and covariances - the parameter covariance matrix.

We will use a quicker (but not as precise) method to estimate the variances and covariances of the estimated parameter values (for example, the scale factor "sf" and slope "c" of a Weibull distribution). We use a formula derived by assuming variations in the data. In order to do this we again will need weights for the data.

We wish to form an estimate for the covariance matrix of the estimated parameters. This estimate will be a function of the objective function, the assumed statistical distribution, the data, and, not so obviously, the estimation algorithm.

The expression we will use comes from Bard [1974] and is derived in Appendix C. It is

$$\mathbf{V_\theta} \approx \mathbf{H^*}^{-1} \left(\partial^2\mathbf{\Phi} / \partial\mathbf{\theta}\, \partial\mathbf{w}\right) \mathbf{V_w} \left(\partial^2\mathbf{\Phi} / \partial\mathbf{\theta}\, \partial\mathbf{w}\right)^{\mathbf{T}} \mathbf{H^*}^{-1} \tag{5-16}$$

where

$\mathbf{V_\theta}$	<u>parameter</u> covariance matrix [eq (1-2)] [dimensions (m x m)]
$\mathbf{\Phi}$	objective function [dimensions (1 x 1)]
$\mathbf{\theta}$	parameter vector (here, sf and c) [dimensions (m x 1)]
$\mathbf{w}$	data vector [dimensions (n x 1)]
$\mathbf{V_w}$	<u>data</u> covariance matrix [dimensions (n x n)]
$\mathbf{H^*} = \partial^2\mathbf{\Phi} / \partial\mathbf{\theta}^2\big\|_{\mathbf{\theta}=\mathbf{\theta}^*}$	[dimensions (m x m)]
*	indicates evaluation at the solution point

If we have "n" data points and "m" parameters to be estimated then eq (5-16) has dimensions (m x m) = (m x m) (m x n) (n x n) (n x m) (m x m).

We will make the assumption that the data covariance matrix is known and that its value is given by the covariance matrix of the order statistics. This is consistent with the assumption previously made, namely, that the weights are functions of the particular parameter values forming the estimate. In fact, we use an explicit value for the data covariance matrix corresponding to the solution values of "sf", and "c".

It must be emphasized that the parameter variance and covariance estimates are just that: each of the parameters has a sampling distribution for its expected value and its variance, making

---

42 Tolerance limits refer to limits found from the actual distribution. Confidence limits refer to limits found from some statistic of the distribution. For example, the mean is a statistic. Percentage points of the distribution of the mean are confidence limits. The ultimate objective is to find tolerance limits.

for four unknown sampling distributions. The values computed may be thought of as samples drawn from the sampling distributions. We are making the (implicit) assumption that the expected values (the bias) of the parameter sampling distributions is small for the problems which we are considering.[43] This assumption needs to be checked by Monte Carlo methods since an analytical formulation is not available. Monte Carlo methods will also provide estimates of the sampling distribution of the parameter variances and covariances.

Applying eq (5-16) to the estimates found using section 5.8.5 and 5.8.6 methods result in the parameter covariance and correlation matrices, and standard deviations given in Table 5-8.

Table 5-8. Parameter Covariance Matrices Using Different Weighting Methods, n = 9

using wts = 1/var

$$\begin{bmatrix} 2.86434 * 10^6 & -238.117 \\ -238.117 & 0.56718 \end{bmatrix}$$

using wts = (covariance matrix)$^{-1}$

$$\begin{bmatrix} 1.18407 * 10^7 & 595.125 \\ 595.125 & 1.12783 \end{bmatrix}$$

The parameter correlation matrices [eq (1-3)] are:

using wts = 1/var

$$\begin{bmatrix} 1 & -0.186817 \\ -0.186817 & 1 \end{bmatrix}$$

using wts = (covariance matrix)$^{-1}$

$$\begin{bmatrix} 1 & 0.162853 \\ 0.162853 & 1 \end{bmatrix}$$

The standard deviations are [sqrt(main diagonal terms)]; example: sqrt($2.86434*10^6$ = 1692:

	using wts = 1/var		using wts = (covariance matrix)$^{-1}$
sf	1692	sf	3441
c	0.753	c	1.062

The estimated parameter values [Table 5-7 methods d and e] are:

	using wts = 1/var		using wts = (covariance matrix)$^{-1}$
sf	31300	sf	31600
c	3.12	c	3.21

The ratios (st dev) / (parameter value)

	using wts = 1/var		using wts = (covariance matrix)$^{-1}$
sf	0.05	sf	0.11
c	0.24	c	0.33

---

43 That this assumption is not necessarily valid for all parameter values is shown in Abernethy, et al, [1983, appendix F].

Weighting using the full covariance matrix, while more tedious computationally, shows that weighting using only the variances is <u>not conservative</u> for this case.

The ratios (sd dev) / (parameter value) show that the scale factor "sf" is much better determined than the slope "c".

We have only nine data points - not too much should be read into these results.

Another set of data of sample size 24 had its parameters estimated using section 5.8.5 and 5.8.6 methods (the same techniques as were used for the set of nine points). The resulting parameter covariance and correlation matrices, and standard deviations are given in Table 5-9.

Table 5-9. Parameter Covariance Matrices Using Different Weighting Methods, n = 24

using wts = 1/var

$$\begin{bmatrix} 435381 & -11.7537 \\ -11.7537 & 0.0582752 \end{bmatrix}$$

using wts = (covariance matrix)$^{-1}$

$$\begin{bmatrix} 3.83218 * 10^{6} & 254.769 \\ 254.769 & 0.324408 \end{bmatrix}$$

The parameter correlation matrices [eq (1-3)] are:

using wts = 1/var

$$\begin{bmatrix} 1 & -0.0737898 \\ -0.0737898 & 1 \end{bmatrix}$$

using wts = (covariance matrix)$^{-1}$

$$\begin{bmatrix} 1 & 0.228496 \\ 0.228496 & 1 \end{bmatrix}$$

The standard deviations are [sqrt(main diagonal terms)]; example: sqrt(435381 = 660:

	using wts = 1/var	using wts = (covariance matrix)$^{-1}$
sf	660	1958
c	0.241	0.570

The estimated parameter values are:

	using wts = 1/var	using wts = (covariance matrix)$^{-1}$
sf	34000	31800
c	3.32	3.47

The ratios (st dev) / (parameter value):

	using wts = 1/var	using wts = (covariance matrix)$^{-1}$
sf	0.02	0.06
c	0.07	0.16

Weighting using the full covariance matrix, while more tedious computationally, shows that weighting using only the variances is <u>not conservative</u> for this case.

The ratios (sd dev) / (parameter value) show that the scale factor "sf" is much better determined than the slope "c". These ratios are smaller than those estimated using nine data points. The slopes and scale factors are comparable.

We have 24 data points - again, not too much should be read into these results. Another sample of 24 points will very likely give different values.

5.10 Uncertainty Estimates

We would like to know how much our computed values would vary due to uncertainty in our estimated parameter values. If we were to take many samples and estimate parameters, how repeatable would our computed values be? For example, the Weibull distribution may be solved for x to obtain

$$x = x_0 + sf * \left(\left\{ \ln[1/(1-P)] \right\}^{1/c} \right) \quad \text{(B-10)}$$

For fixed P, what is the uncertainty in x due to variability in x_0 , sf, and c? We address this question by deriving an expression for the variance of x.[44] If we expand eq (B-10) about x in a Taylor's series, keep only the first terms, and assume that we have many samples we obtain

$$\sigma_{\mathbf{x}}^2 = \mathbf{dvec}^{\mathbf{T}} * \mathbf{pac} * \mathbf{dvec} \quad \text{(5-17)}$$

where

dvec	column vector of derivatives with respect to parameters
pac	parameter covariance matrix

Eq (5-17) was derived under the further assumption that we have no bias (eq 5-2).

Using eq (B-10) we obtain

$$\sigma_x^2 = \begin{bmatrix} \partial x/\partial sf & \partial x/\partial c & \partial x/\partial x_0 \end{bmatrix} \begin{bmatrix} \sigma_{sf}^2 & cov(sf,c) & cov(sf,x_0) \\ cov(c,sf) & \sigma_c^2 & cov(c,x_0) \\ cov(x_0,sf) & cov(x_0,c) & \sigma_{x0}^2 \end{bmatrix} \begin{bmatrix} \partial x/\partial sf \\ \partial x/\partial c \\ \partial x/\partial x_0 \end{bmatrix} \quad \text{(5-18)}$$

The parameter covariance matrix **pac** is always symmetric.

For the examples in the previous section, x_0 was fixed, so "sf" and "c" are the only variables so Eq (5-18) becomes

---

44 This is the third variance we work with. One variance is that of the entire distribution [eq (4-5d)], the second is the set of variances, one for each order number (section 4.5.2), while the third is the variance of a fitted or computed variable such as "x" above.

$$\sigma_x^2 = \begin{bmatrix} \partial x/\partial sf & \partial x/\partial c \end{bmatrix} \begin{bmatrix} \sigma_{sf}^2 & cov(sf,c) \\ cov(c,sf) & \sigma_c^2 \end{bmatrix} \begin{bmatrix} \partial x/\partial sf \\ \partial x/\partial c \end{bmatrix} \tag{5-18a}$$

Example 1: n = 9

From Table 5-8, right-hand column, we have $\mathbf{pac} = \begin{bmatrix} 1.18407*10^7 & 595.125 \\ 595.125 & 1.12783 \end{bmatrix}$.

pac stays constant as long as the values of "sf" and "c" don't change.

For P = 1 - 1/e = 0.632 (the characteristic value) we have

$$\sigma_{x\ \text{for}\ P=1-1/e}^2 = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 1.18407*10^7 & 595.125 \\ 595.125 & 1.12783 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = 1.18407*10^7$$

so that $\sigma_{x\ \text{for}\ P=1-1/e} = 3441$ with x = 31585. The ratio $\sigma_x/x = 3441/31585 = 0.109$.

When P = 0.9 (corresponding to the probability for the largest in the sample) we have

$$\sigma_{x\ \text{for}\ P=0.9}^2 = \begin{bmatrix} 1.29687 & -3318.81 \end{bmatrix} \begin{bmatrix} 1.18407*10^7 & 595.125 \\ 595.125 & 1.12783 \end{bmatrix} \begin{bmatrix} 1.29687 \\ -3318.81 \end{bmatrix} = 2.72142*10^7$$

so that $\sigma_{x\ \text{for}\ P=0.9} = 5217$ with x = 40962. The ratio $\sigma_x/x = 5217/40962 = 0.127$.

Not only is the standard deviation increasing with P; the ratio sd / x is also increasing.

Example 2: n = 24

From Table 5-9, right-hand column, we have $\mathbf{pac} = \begin{bmatrix} 3.83218*10^6 & 254.769 \\ 254.769 & 0.324408 \end{bmatrix}$.

pac stays constant as long as the values of "sf" and "c" don't change.

For P = 1 - 1/e = 0.632 (the characteristic value) we have

$$\sigma_{x\ \text{for}\ P=1-1/e}^2 = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} 3.83218*10^6 & 254.769 \\ 254.769 & 0.324408 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = 3.83218*10^6$$

so that $\sigma_{x\ \text{for}\ P=1-1/e} = 1958$ with x = 31839. The ratio $\sigma_x/x = 1958/31839 = 0.061$.

When P = 0.96 (corresponding to the probability for the largest in the sample) we have

$$\sigma^2_{x \text{ for } P=0.96} = \begin{bmatrix}1.40069 & -4331.56\end{bmatrix} \begin{bmatrix} 3.83218*10^6 & 254.769 \\ 254.769 & 0.324408 \end{bmatrix} \begin{bmatrix} 1.40069 \\ -4331.56 \end{bmatrix} = 1.05137*10^7$$

so that $\sigma_{x \text{ for } P=0.96} = 3242$ with $x = 44597$. The ratio $\sigma_x/x = 3242/44597 = 0.073$.

Again, not only is the standard deviation increasing with P; the ratio sd / x is also increasing.

The scale factors and slopes are comparable for the two examples (31600, 3.21) and (31800, 3.47). The square root of the ratio of sample sizes = sqrt(24/9) = 1.6. The ratios of (sd/value) at the characteristic values is (0.109/0.061) = 1.8. It seems reasonable that increasing sample size should reduce the uncertainty. While these numbers are suggestive, the relation of change of sd/value ratio with sample size requires much more work to be established, and so a reduction of uncertainty with $n^{-1/2}$ is not yet established.

5.11 Comments on "Goodness" of Fit, Confidence Intervals, Tolerance Intervals and Limits

"Goodness" of fit deals with such topics as the size and distribution of residuals, indifference regions (how much may the solution be varied before there is a significant difference in the estimated parameter values), how well are parameters determined (in section 5.9, the scale factor "sf" was better determined than the slope "c" if only the parameter covariance standard deviation is used).

Confidence intervals deal with the probability that the true value of a statistic lies within a given interval. In the parameter covariance estimates made in section 5.9 we implicitly assumed that the parameter values for "sf" and "c" were estimates of the true (but unknown) values of the means of the sampling distribution of scale factors and slopes. Specific numerical values depend upon the distribution. It is quite likely that, for the nonlinear problems which we deal with here, that the distributions of some of the parameters of interest are not symmetric. In particular, we should not assume normality without some justification.

Tolerance intervals deal with questions of the following form: "for xx confidence that yy of the population lies between standard deviation limits the tolerance interval factor is ±zz on the standard deviation. An example: "The interval that would contain 90% (yy) of the parent population with 95% (xx) is ±4.2 * (sample standard deviation) [here zz = 4.2]." Specific numerical values depend upon the distribution. As before, it is quite likely that, for the nonlinear problems which we deal with here, that the distributions of some of the parameters of interest are not symmetric. In particular, we should not assume normality without some justification.

Tolerance limits imply that we have the actual distribution of the parameter. Then we can read off the pertinent numbers. The distribution does not have to be symmetric. In order to get the distribution, resort must usually be made to a great deal of simulation so that the histogram of number of simulations forms a good approximation to the population distribution. When we use tolerance limits, we are no longer dealing with something calculated from a sample (a statistic), but with a value from the population. We would always like to do this, but can rarely can do so due to schedule and budget constraints.

5.12 Parameter Estimation Summary

We developed a conceptual framework within which to do parameter estimation. This framework is always present whenever parameters are being estimated where uncertainty is present. The uncertainties arise from four, possibly five sources:

a) measurement uncertainty,
b) sample size effects,
c) the use of different models, and
d) different estimation procedures.
e) for statistical distributions: method for assignment of probability [P = r/(n+1) etc.].

This conceptual framework applies to any fitting function where we estimate parameter values no matter whether the fitting function is a statistical distribution or not, or whether we are engaged in linear or nonlinear parameter estimation.

We developed this framework by considering uncertainties and errors, the sampling distribution, the form of an appropriate error model, and a weighted objective function.

When the fitting function is a probability distribution we assigned probabilities to the sample points.

Examples using a Weibull distribution using five different fitting methods (five different objective functions) were shown.

Results from two of the methods ("d" and "e") were used to form estimates of the parameter covariance matrix. Use of the full data covariance matrix showed that the use of only the data variance matrix resulted in a non-conservative parameter covariance estimate.

The uncertainties for a given sample size were not constant, but depend upon the probability value.

Estimates of the uncertainties in the values of "x" (such as bending moment) computed using the estimated parameter values were shown for different sample sizes. The larger sample size had smaller uncertainties.

6.0 REPRESENTATION OF COMBINED VERTICAL AND LATERAL BENDING

6.1 Theory

For both backspline and PVC ship models the moments are measured using strain gage readings which are calibrated with known moments. These physical ship models, especially the backspline ship models, are strong enough that the measured moments are linear with respect to the gage readings (The ship model does not enter the non-linear range: i.e., no permanent set occurs while testing.)

The axis system, ***fixed to the ship***, is shown below. *It must be strongly emphasized that the axis system in which the moments are measured moves with the ship.* If the ship rolls 30 degrees, so do the axes.

Vertical bending moments cause rotation about the x-x axis, while lateral bending moments cause rotation about the y-y axis. At any time there is a vertical moment magnitude labeled *Vert bm* and a lateral bending moment magnitude labeled *Lat bm*. The resultant moment, in ship coordinates, is labeled *combined moment*, and the axis about which it rotates is labeled *combined moment axis*. The angle from the vertical bending axis is labeled *moment axis angle*.

The magnitude of the combined moment and the moment axis angle change as a function of time. (The combined moment vector changes in both magnitude and direction.)

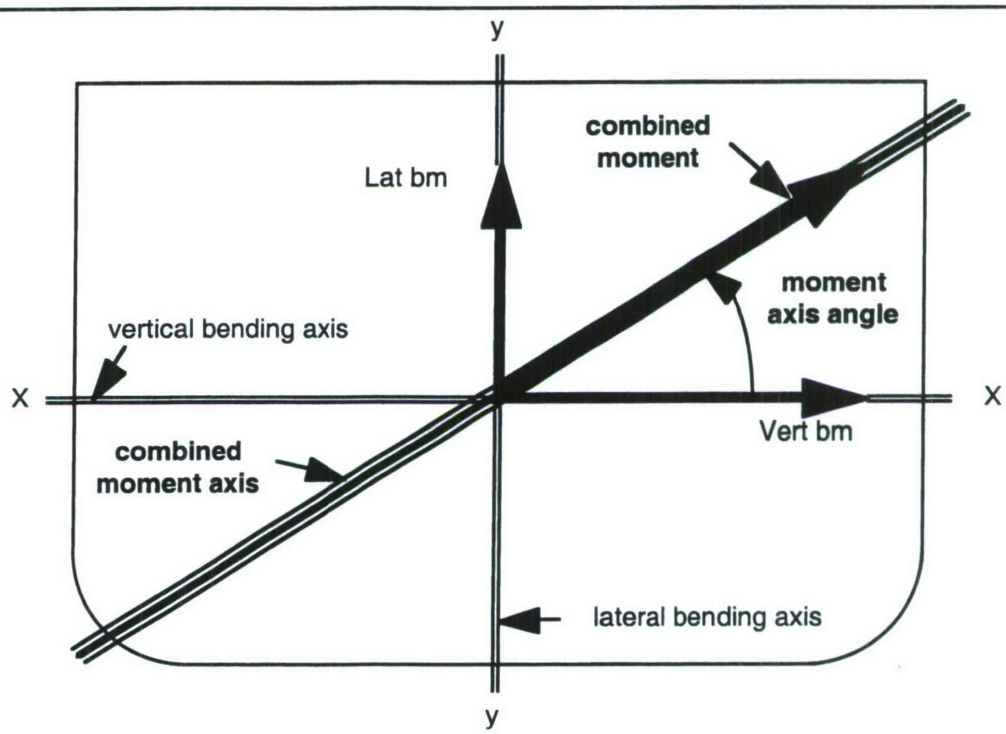


Figure 6-1. Model for Collecting Moment Data

The combined moment is the vector sum of "Vert bm" and "Lat bm". Knowledge of this magnitude and its associated moment axis angle allows us to find the components of the combined moment in any coordinate system.

In Figure 6-2 the magnitude and orientation of the combined moment vector remains unchanged with respect to the geometric cross section. The outline of the hull section remains the same, else a different time history would result. The model internal structure is replaced by the actual structure when estimating stresses.

For finding stresses this combined moment is applied at the intersection of the principal axes.[45] It does not matter that the principal axes may not be at the location of the original data collection principal axes from which the magnitude and direction of the combined moment and moment axis angle were computed. In Figure 6-2 the black square shows the original position of the original data collection principal axes. In general, the principal axes for estimating stress are both shifted and rotated from the original position of the data collection principal axes.

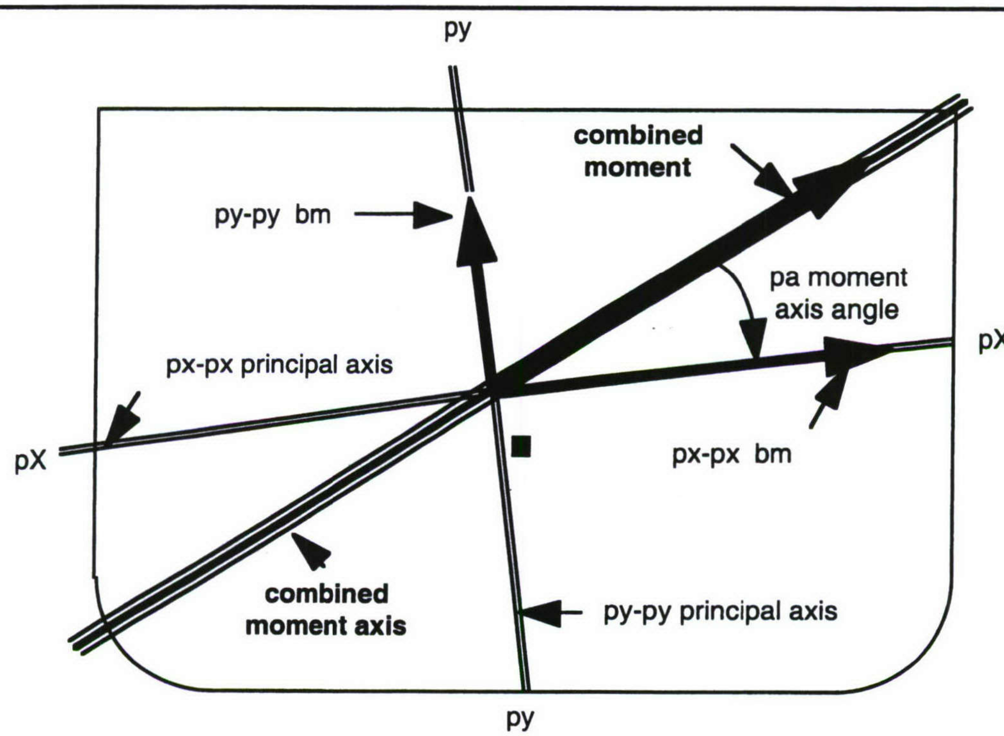


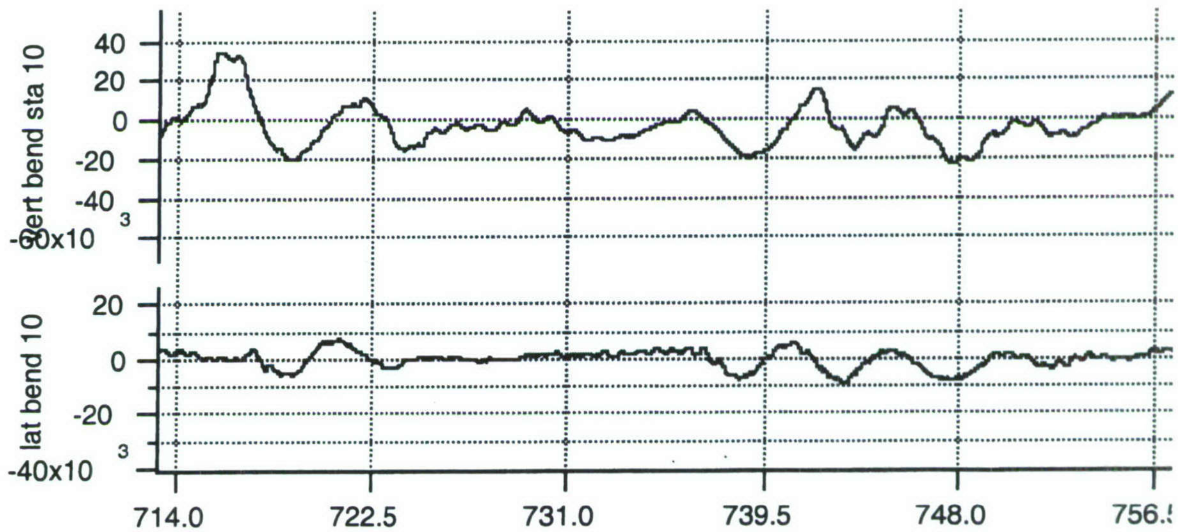
Figure 6-2. Model for Using Moment Data

45 The principal axes are defined as the point about which the forces and moments acting on the cross section are in balance. In addition, there cannot be any net moment normal to the applied moment. It can be shown that these conditions are sufficient to uniquely determine the location and orientation of the principal axes. These conditions apply whether the stress-strain relationship is linear or non-linear (it is assumed that plane sections remain plane).

The combined moment and moment axis angle time history remains invariant. Consequently changes in the ship structure result in changes in the stress field, but do not affect the combined moment and moment axis angle time history.

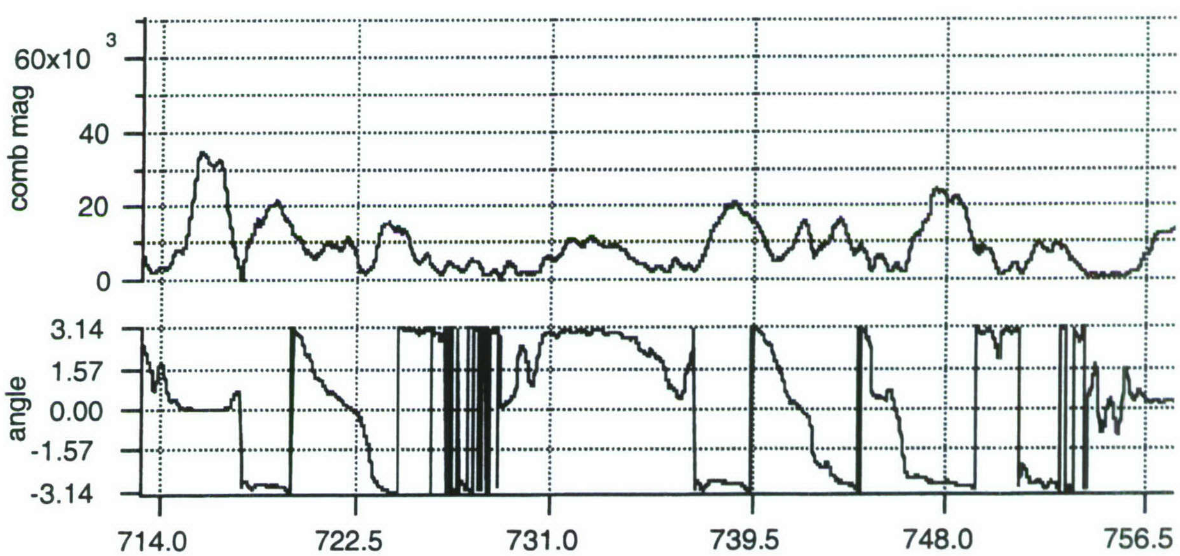
6.2 Ship A Data Set

A portion of a vertical and lateral bending time history for Ship A ***full-scale*** is shown in Figure 6-3. The corresponding combined vertical and lateral bending time history is shown in Figure 6-4.



(Vertical scale (bending) is in *m-mt* units; horizontal scale is time in *seconds*.)

Figure 6-3. Portion of Vertical and Lateral Bending Time History



(Lower vertical scale is in *radians*; horizontal scale is time in *seconds*.)

Figure 6-4. Portion of Combined Vertical and Lateral Bending Time History

The top trace in Figure 6-4 is magnitude = (vertical^2 + lateral^2)$^{1/2}$. The bottom trace in Figure 6-4 is the moment axis angle. The moment axis angle is given by maa = $\tan^{-1}$(lateral, vertical). Two arguments are needed since quadrant information needs to be saved. A vertical hog bending moment with no lateral bending moment has a moment axis angle = 0, while a vertical sag bending moment with no lateral bending moment has a moment axis angle = $\pm \pi$ ($\pm$ 3.14...). A time history of moment axis angle for vertical sag shows rapidly changing values since + π and - π are the same angle. The explanation of the vertical bars every 8.5 seconds is given in the next section.

The original values of vertical and lateral bending are recovered by

vertical = magnitude * cos(moment axis angle)
lateral = magnitude * sin(moment axis angle)

6.3 Independence Interval Selection

The theory developed in Sections 3 and 4 for estimating lifetime loads assume that the data points used are statistically independent of each other. How can a time interval be selected such that this criterion is met?

The best way is to select the maximum and minimum combined bending moment in each pitch cycle. Since the data used in this example were collected in such a manner that the pitch record could not be matched to the bending moment record, recourse was made to another method. The method selected was to determine the independence interval using the autocorrelation function of the pitch record. The autocorrelation function of a signal is given by Burrington and May [1970]

$$\varphi(k) = \frac{\lim}{n \rightarrow \inf}[1/(2n+1)] \sum_{j=-n}^{j=n} x_{j+k} * x_j \tag{6-1}$$

where

n 1/2 number of points
k lag: k = 0, 1 , ... , as many as are possible with finite n

The maximum value occurs for k = 0. The autocorrelation function is symmetric about k = 0. The autocorrelation function measures the influence of the j^{th} observation on the $(j+k)^{th}$ observation. To the extent that independence holds, this influence approaches zero. Put another way, to what extent does knowledge of the value at "j" help us to predict the value at "j+k"? If we had swell, the autocorrelation function would not die away; to the extent that it dies away, we have a random process (or a dead calm).

The autocorrelation of the pitch record is shown in Figure 6-5. A strong peak occurs at lag t = 8.5 seconds. This corresponds to sag followed by hog, or hog followed by sag and so corresponds to a fairly strong correlation. It is of no concern to us since we are going to treat hog and sag separately. The next peak occurs at lag t = 17 seconds, and has a magnitude ~ 10% of the max peak at lag t = 0. The conclusion is that the use of an interval = 8.5 seconds will be adequate. This means we read a maximum and minimum value in each 8.5 second interval. For each such value read there is a corresponding moment axis angle which we also record.

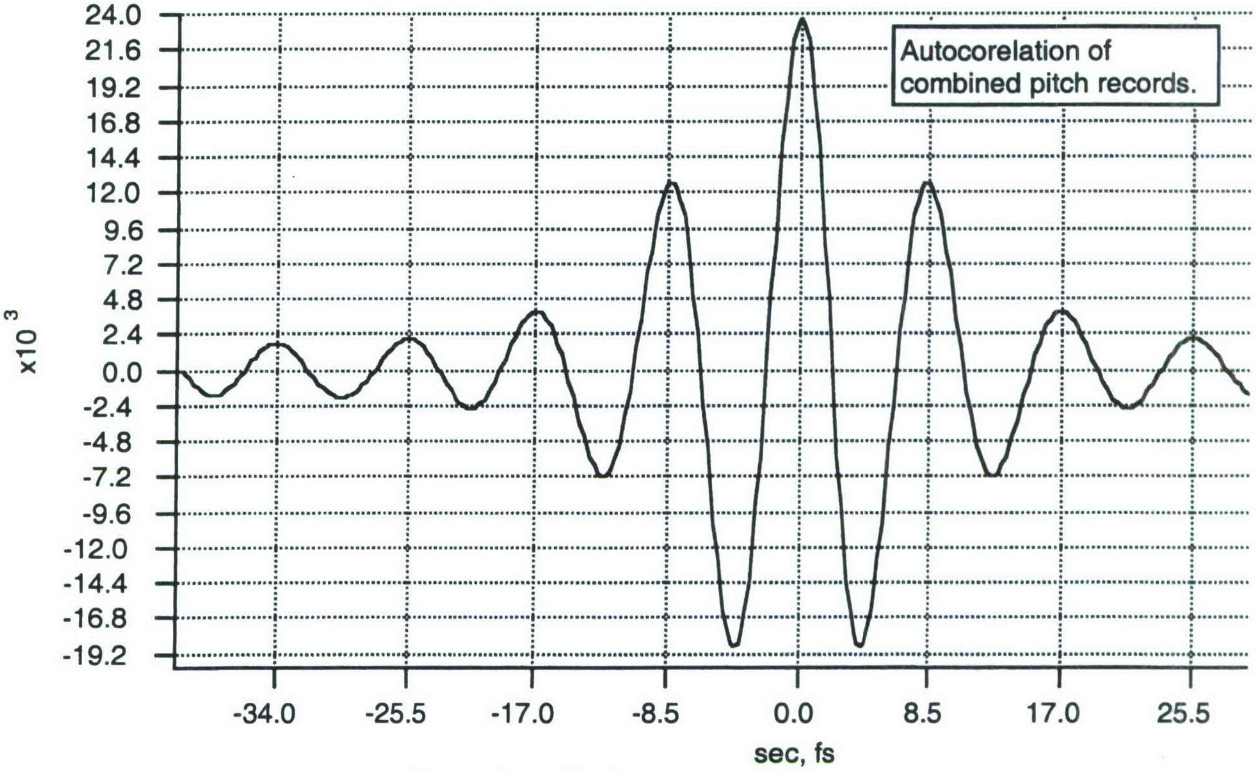


Figure 6-5. Pitch Autocorrelation

How good is the assumption that values read an average of 8.5 seconds apart are independent? Figure 6-6 is an autocorrelation plot of the 75 sag values read in the first 2/5 of the run. The values on the axis correspond to the number of 8.5 second intervals. It will be seen that the autocorrelation function drops to about 15% of its peak value after one interval, so our assumption of independence using an 8.5 second interval appears valid.

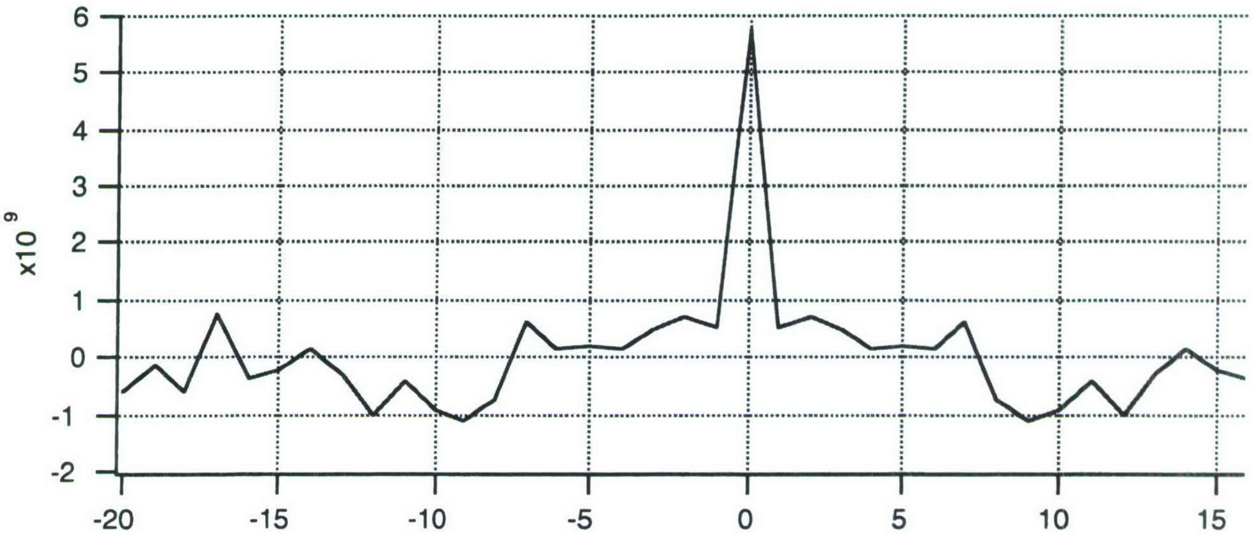
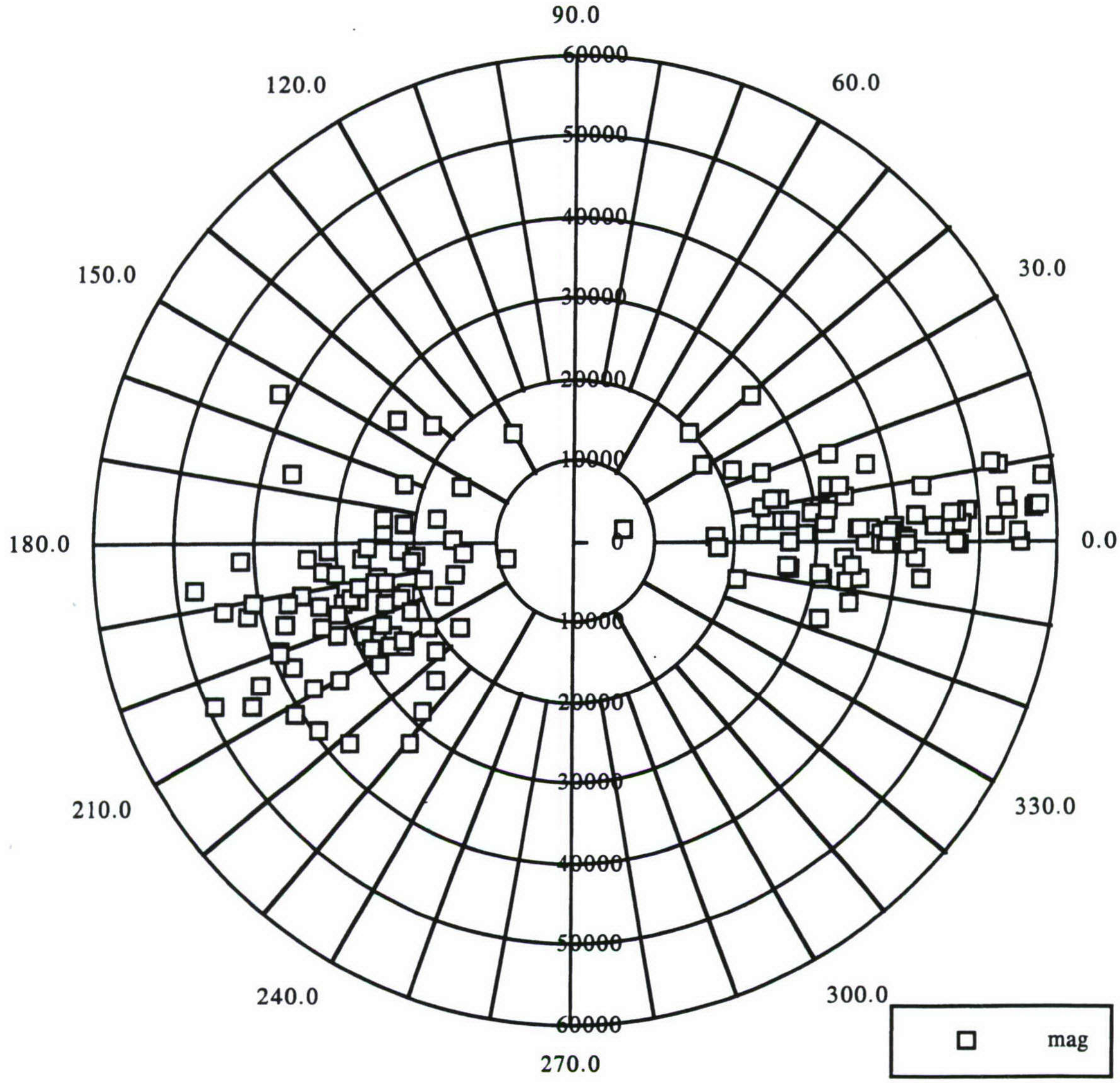


Figure 6-6. Sag Series Autocorrelation

6.4 Representation of Combined Moments

The combined moments are represented on a polar plot as shown in Figure 6-7. The values are from only 2/5 (40%) of the run, and are shown to illustrate the method.



(Combined bending moments are in *m-mt*; angles are in *degrees.*)
(pure hog = 0 degrees, pure sag = 180 degrees)

Figure 6-7. Combined Vertical and Lateral Bending Magnitude Polar Plot

The angles are expressed in degrees rather than in the radians shown in Figure 6-4.

The maximum combined hog magnitude of about 59,000 m-mt is about 20% larger than the maximum combined sag magnitude of about 49,000 m-mt. This is unusual, but, as mentioned above, the data represents only a small portion of the run. The angular spread of the sag values is much greater than the angular spread of the hog values.

Weibull plots of all combined sag and hog events are shown in Figure 6-8 without respect to moment axis angle. The truncation value was set = 0 since it is possible for bending moments to approach zero. "lnln sag" means "lnln(1/(1-P)))" for the sag sample probabilities.

The corresponding Weibull scale factors are sag $= e^{-b/c} = e^{-(-37.873)/3.651} = 32{,}000$ m-mt and hog $= e^{-(-33.516)/3.155} = 41{,}100$ m-mt (Appendix B).

Weibull slopes above 3 are of interest since the usual largest slope is a Rayleigh slope = 2. (Recall that a normal distribution is approximated in Weibull space by a slope ~ 3.46.)

Since neither plot shows deviations uniformly distributed about a straight line, further analysis should be done, especially for hog, since the Weibull plot data shows a tendency to have an "s" shape.

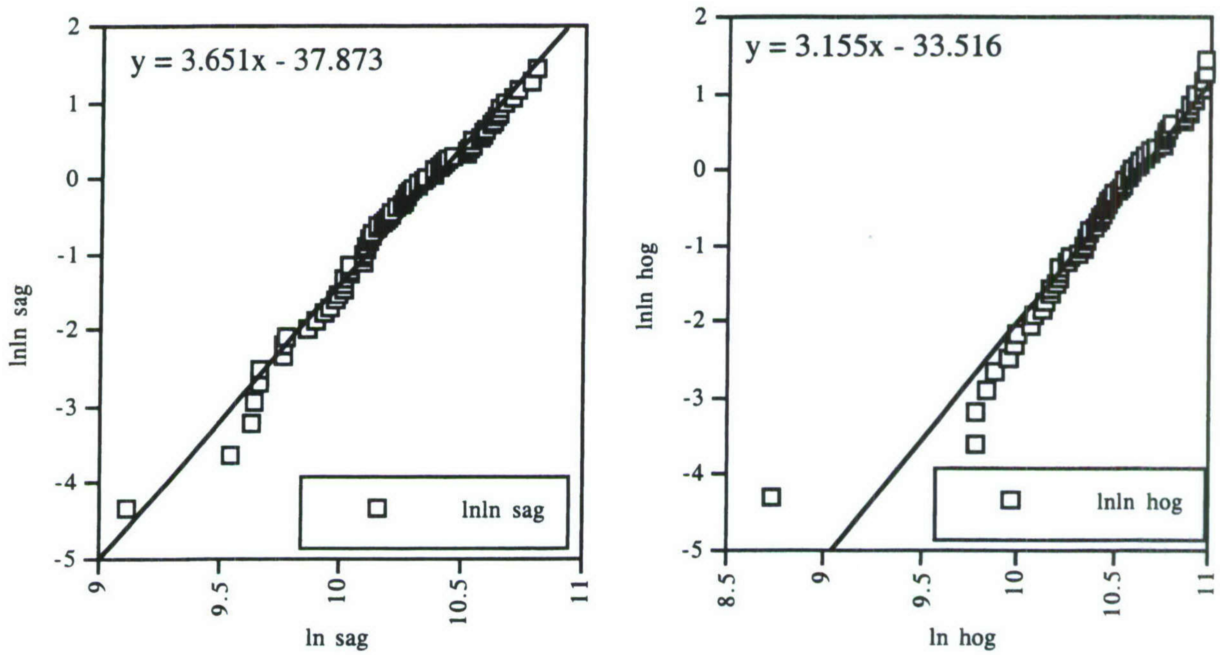


Figure 6-8. Weibull Plots for All Combined Sag and Hog Events

6.5 Sag by Moment Axis Angle

The sag data shown in Figure 6-7 were sorted by 10 degrees zones: 180-189, 190-199, 200-209, and 210-219 degrees. Weibull plots of the data in these 10 degree zones are shown in Figure 6-9. In each zone, two fits are shown: a fit in Weibull space and a fit in original, or data collection space.

While this is not a large amount of data, consisting of 2/5 of a run with heading 30 degrees away from head seas, there is a tendency for the Weibull distribution slopes to increase then decrease with moment axis angle (slopes = 2.9, 3.6, 3.6, and 3.2) as the moment axis central angle changes (185, 195, 205, and 215 degrees).

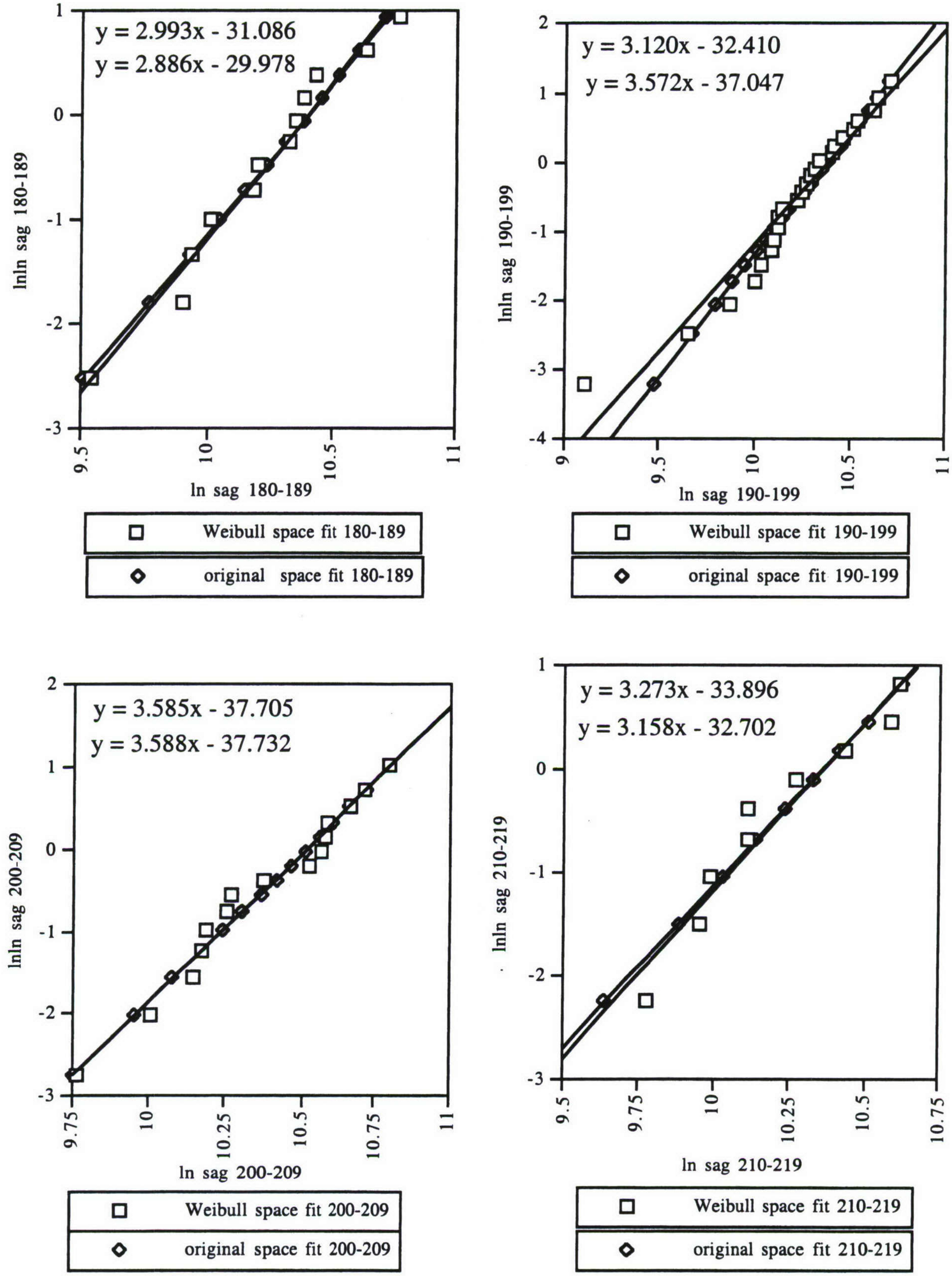


Figure 6-9. Sag Weibull Plots by 10 Degree Zones

The upper equation in each plot shows the fit in Weibull space, while the bottom equation in each plot shows the fit in original, or data collection space.

The data plot better by moment axis angle zone (Figure 6-9) than all combined together (Figure 6-8).

Each moment axis angle zone (180-189, 190-199, 200-209, and 210-219 degrees) corresponds to a cell - conditions within each zone are statistically stationary - the basic requirement defining a cell. For combined vertical and lateral bending each cell corresponds to a plane sector on the polar plot.

Use of Combined Load Cells

An example showing estimation of extreme values for combined vertical and lateral bending using cells is shown in section 9.4. An uncertainty analysis for the extreme value is presented in section 9.5.

6.6 Representation of Combined Vertical, Lateral, and Torsional Bending

Combined vertical, lateral, and torsional bending can be represented using a sphere. The magnitude will be given by radius = $(\text{vertical}^2 + \text{lateral}^2 + \text{torsion}^2)^{1/2}$. It will be necessary to use two angles rather than the single moment axis angle used for combined vertical and lateral bending. The data will plot as three dimensional sectors coming to a point at the origin of the sphere.

The result will again be cells where now a cell is the sector of a sphere.

In this test the magnitude of the torsional component is small compared to vertical and lateral bending so this combined analysis does not have to be performed.

6.7 Summary

A method for representing combined vertical, lateral, and torsional bending at all instants of time, consistent with the cell method, has been presented. It has been illustrated for the combined vertical and lateral case.

The plots did not distinguish wave and whip components - the combined wave plus whip signal was used for either vertical or lateral bending.

How to select an independence interval has been illustrated.

7.0 COMBINED WAVE PLUS WHIPPING DISTRIBUTIONS

The preliminary look at the Ship A data in section 6 showed that the combined wave plus whipping bending moment for the test condition analyzed could reasonably be characterized using Weibull distributions for each vertical plus lateral zone (cell). A more general case occurs when the combined wave plus whipping bending moment data for either vertical or lateral bending (or both) cannot be reasonably described using a single distribution, or by a single distribution having constant parameters.

Two examples will be shown to illustrate methods of approach in these cases. Both are ship models tested running into random head seas so that lateral bending is very small. The first is designated as Ship B, the second is designated Ship C.

The term "low-pass" (LP or lp) refers to the continuous wave induced component of a signal such as bending moment, the term "high-pass" (HP or hp) refers to the whipping induced component of a signal. These terms are used because the two domains have a wide frequency separation (full-scale wave periods which cause significant response are on the order of several seconds, while the full-scale first mode whipping period is on the order of a second.

It is emphasized that we are dealing with the response to the (unknown) load(s). We do not know, nor do we usually care, what the load is. The ship is designed based on the response to the loads, not the loads. In short, we build and test response models, not load models,[46] even though they are described as loads models.`

7.1 Summary of the Method(s)

The following procedure(s) are used:

1) Develop a polar plot of combined vertical and lateral bending;
2) Select a sector of this polar plot for analysis;
3) Using the data in the chosen sector, sort the collection of vertical and lateral bending by magnitude and try to find a distribution which fits. If a distribution fits, we are finished (this is the case for hog).
   If no distribution fits:
4) For either vertical or lateral bending:
   a) make a plot of low-pass bending;
   b) fit the low-pass component
      For the Ship B low-pass data, one distribution worked;
      For Ship C, two distributions were needed to fit the low-pass data.

The high-pass component values cannot be fitted using a single distribution unless the high-pass values are statistically independent of the low-pass component values - a situation which did not occur with either the Ship B or Ship C data sets (and is congruent with past experience: larger lp values tend to be associated with larger hp values). Put another way, *the*

[46] This alternative nomenclature was proposed years ago. However, the loads community continues to refer these types of tests as "loads model" tests.

high-pass component values are usually correlated with the low-pass component values. This means that the high-pass component values are usually conditional upon the low-pass component values.

This conditional dependence is why the problem of finding the distribution function of t = lp + hp cannot be found by convolution: the convolution operation results from the assumption that the distribution of the lp component is independent of the distribution of the hp component.

Continuing on:

c) make a plot of high-pass vs. low-pass bending;

d) divide the low-pass values into zones;

i) for each low-pass zone, fit distributions to the high-pass values.

For Ship B, the high-pass components were satisfactorily fitted using Weibull distributions.

For Ship C, the high-pass components were satisfactorily fitted using log normal distributions (Weibull distributions failed).

ii) if all the low-pass distributions are of the same type, fit parameter value functions to the high-pass parameter distribution values such as scale factor and slope for Weibull distributions, or mean and standard deviation for the log normal distribution.

For Ship B two different sets of Weibull parameter functions were needed; for Ship C one log normal parameter function set sufficed.

The values of the parameter functions depend upon the values of the low-pass component: this is where the conditional dependence of hp upon lp is expressed.

iii) If the above technique for finding parameter fitting functions fails, then recourse must be made to fitting piecewise continuous functions to the hp components. (This is not necessary for the examples shown in this report.)

7.2 Ship B

The Ship B data are ***model scale*** vertical bending in ***in-lb*** units. These data were worked up before the development of the method for combining vertical and lateral bending. The time of occurrence and magnitude of combined peaks are given. The time of occurrence and magnitude of the nearest low-pass and high-pass peak are given. This means there is some mismatch between the value of the combined peak and the sum of the nearest low-pass and high-pass peak.[47] For this data set, it is necessary to work with the magnitudes of the nearest low-pass and high-pass peak.

[47] The values in this data set are close to the values obtained if a combined peak was split into the lp and hp components occurring at the same time as the combined peak.

Not every occurrence of a large wave sag value results in a whip. Of the 1782 independent low-pass wave events 152 of these had associated whips for an overall frequency of whipping = 8.5%.

Figure 7-1 is a plot of all 152 whip LP combined sags vs. HP combined sags. The trend is that larger LP combined sags have larger associated HP combined sags. The HP combined sag splits into two domains at a LP combined sag value of about 16,000 (16k) in-lb.

"HPcombmin" is the Ship B test nomenclature for the total vertical bending high-pass component. "LPcombmin" is the Ship B test nomenclature for the total vertical bending low-pass component.

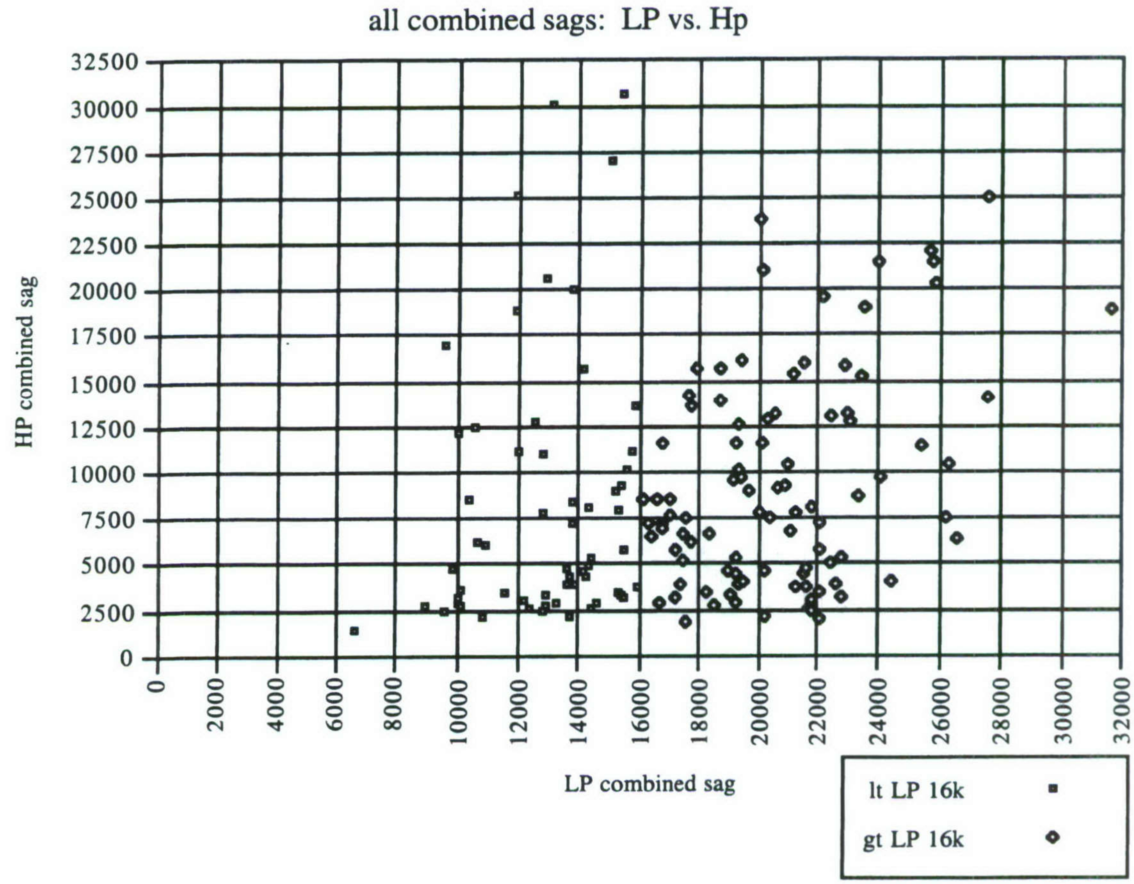


Figure 7-1. Set of 152 Combined Sags: LP vs. HP, in-lb units

There is a lower value (about 6,700 in-lb) of LP combined sag below which no whips occurred. If one looks at the largest value of LP combined sag for which a whip occurs, there is only one occurrence in the LP combined sag zone from 28,000 to 32,000 in-lb. While 32,000 in-lb should not be taken as an upper limit beyond which no whips occur, it does suggest that the frequency of whipping is decreasing rapidly once the LP combined sag value is greater than about 28,000 in-lb.

7.2.1 Fit Distribution(s) to the Low-pass Component(s) and the High-pass Component(s)

None of the solutions shown in this section (section 7) have been obtained using weighting by data variances. Weighting should be used when the final fits are evaluated. This can be computationally intensive (section 5).

7.2.1.1 Fit Low-pass Component(s)

Although it seems reasonable, the Weibull distribution when no threshold value is used ($x_0 = 0$) does a poor job of estimating large values. It under estimates the probability of occurrence of large magnitude, low-pass events, and so is not conservative.

When a threshold is used, the Weibull distribution does a good job of fitting the low-pass component. The meaning of this threshold is difficult to explain. A start is to note that no experimentally observed combined wave plus whip events occurred below a LP value of about 6,700 in-lb. The threshold then can be interpreted to mean that no whip events occur below a LPmin value of 4,369 in-lb. When a three parameter Weibull distribution is used, the best fit in original space is obtained for $x_0 = 4{,}369$ in-lb, scale factor sf = 14,919 in-lb, and slope c = 3.048. Figure 7-2 is the Weibull plot.

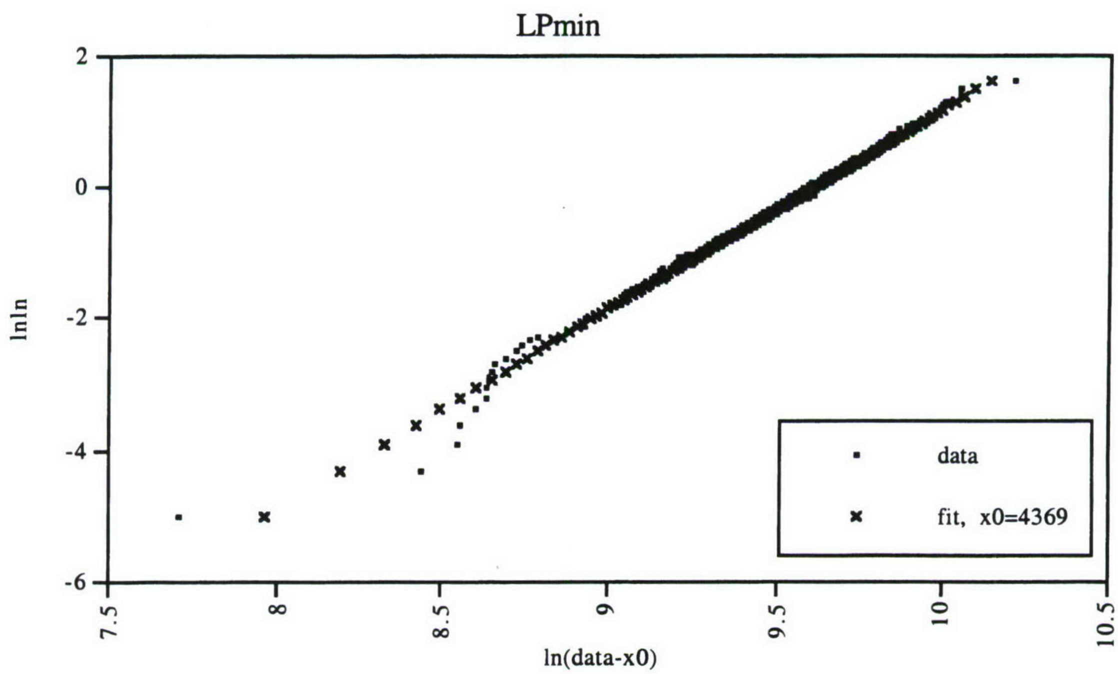


Figure 7-2. Weibull Fit For LP Component

7.2.1.2 Fit High-pass Component(s)

All the high-pass components can be reasonably fitted using a Weibull distribution.

Each of the Weibull plots in Figures 7-3 and 7-4 is for a 2,000 (2k) in-lb LP zone. Figure 7-3 shows Weibull fits for lp region 1 (6k-16k), while Figure 7-4 shows the Weibull fits for lp region 2 (16k-28k). For example, Figure 7-3a plots all the high-pass values found in LP zone 8k-10k.

There are two equations given in each plot. The upper equation is the result of Weibull fitting in Weibull space {[(lnln(1/(1-P)) vs. ln(data)] or [(ln(data) vs. lnln(1/(1-P))][48]}. The first pair, with ln(data) as the independent variable assumed to have no error, was used. This correspond to the label "data" in the plots.

The lower equation in each plot is the result of fitting in original space[49] [(data vs. P) or (P vs. data)]. The first pair, with $P = r/(n+1)$ as the independent variable assumed to have no error, was used (section 5.8.4). The marked superiority of fitting in original space is especially apparent in Figures 7-3c and 7-3d where the distortion due to using logs is obvious.

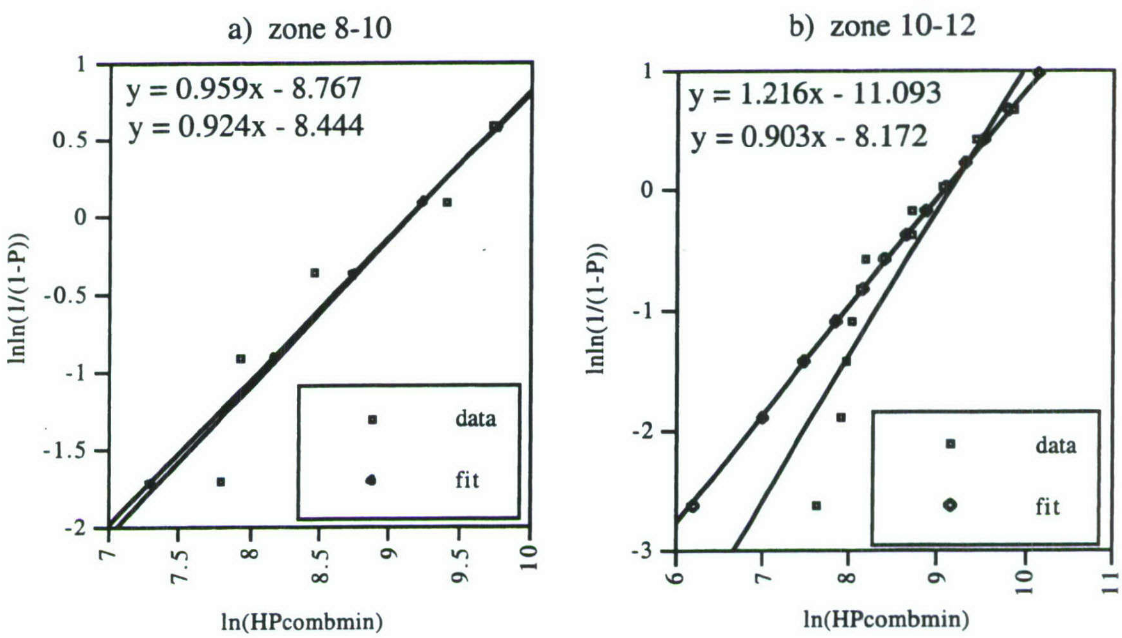


Figure 7-3. Hp Fits for Region 1 LP Combined Min 6k - 16k

48 This is the preferred method since the variability in sample size is apt to be less than the variability in magnitudes. The first pair gives a "quick look" estimate.

49 "Original space" means using non-transformed data values sorted in increasing order of magnitude. In $P = r/(n+1)$ r is the order number (r = 1, smallest; r = n, largest), and n is the sample size. The computation is non-linear in x0, sf, and c. It goes quickly on modern spreadsheets.

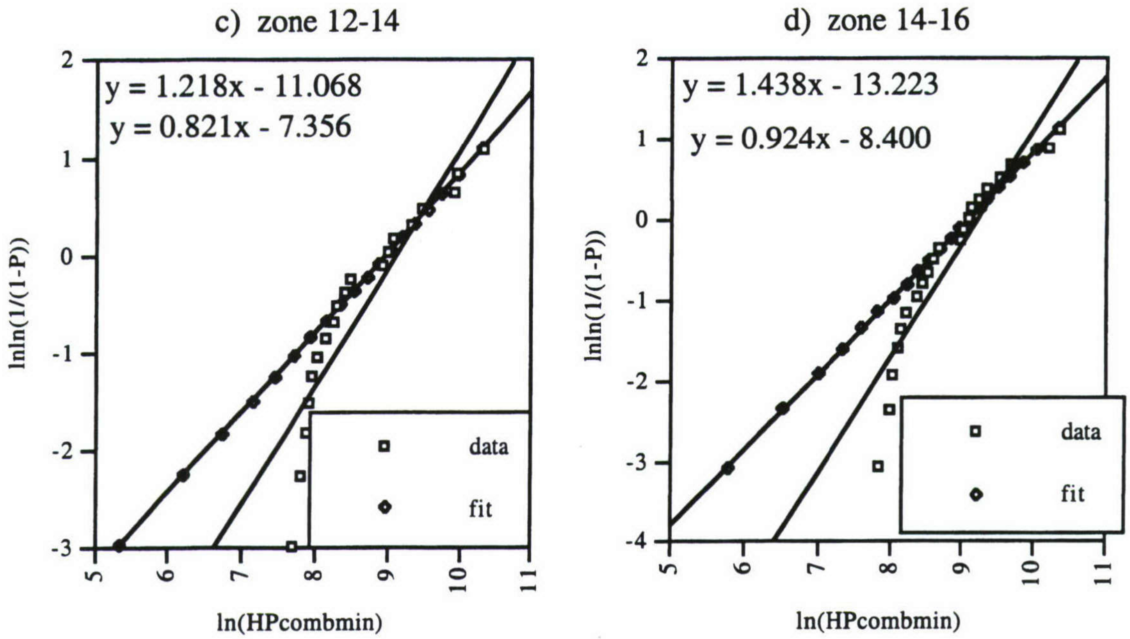


Figure 7-3. Hp Fits for Region 1 LP Combined Min 6k - 16k (continued)

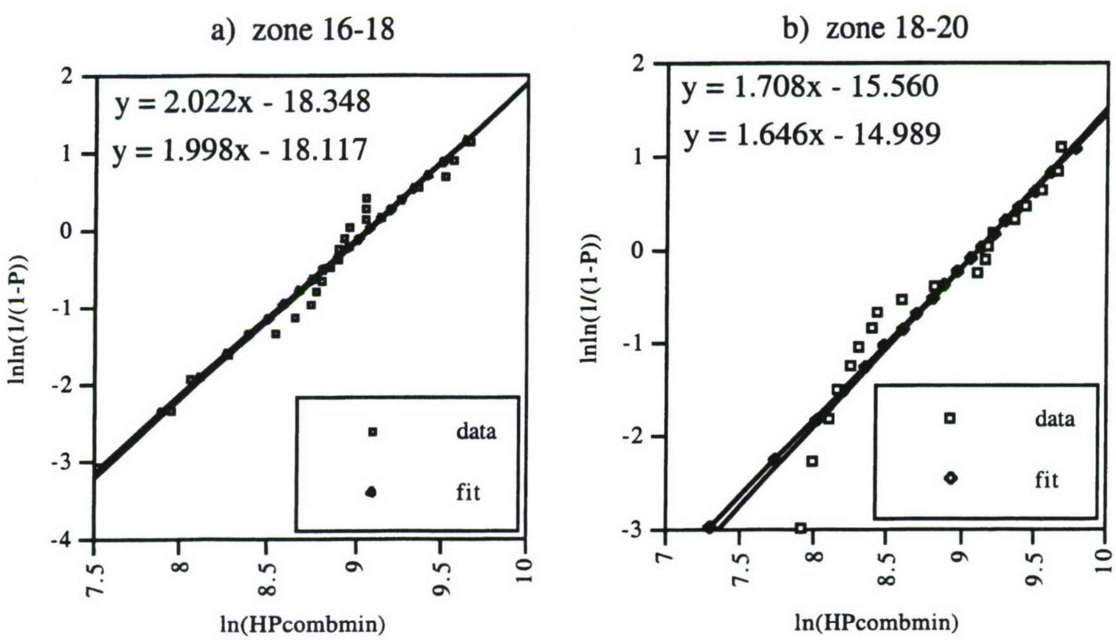


Figure 7-4. Hp Fits for Region 2 LP Combined Min 16k - 28k

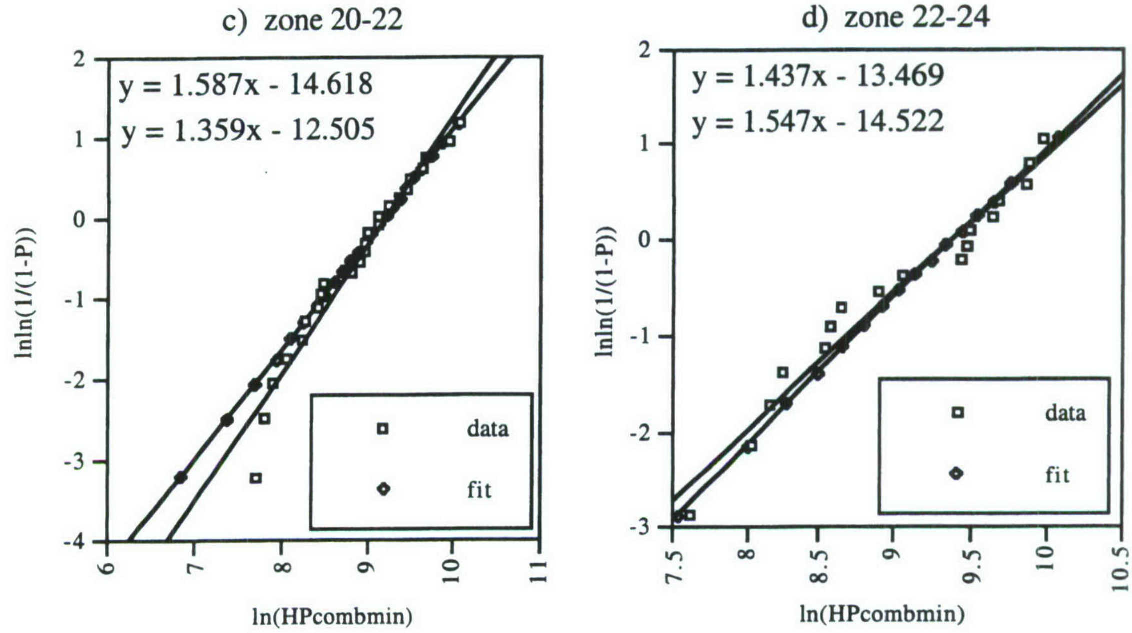


Figure 7-4. Hp Fits for Region 2 LP Combined Min 16k - 28k (continued)

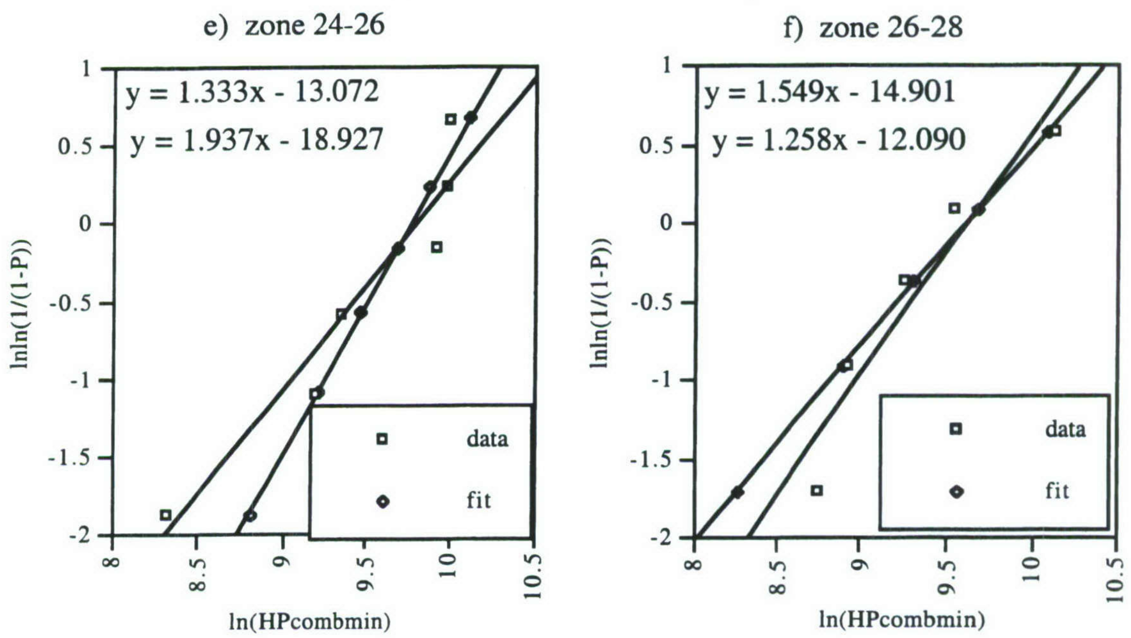


Figure 7-4. Hp Fits for Region 2 LP Combined Min 16k - 28k (continued)

The next step is to form equations for the variation with vb lp for the vb hp slope and the scale factor. There will be two of each since there are different Weibull parameters for the two lp regions. Plots of the slopes and scale factors are shown in Figures 7-5 and 7-6.

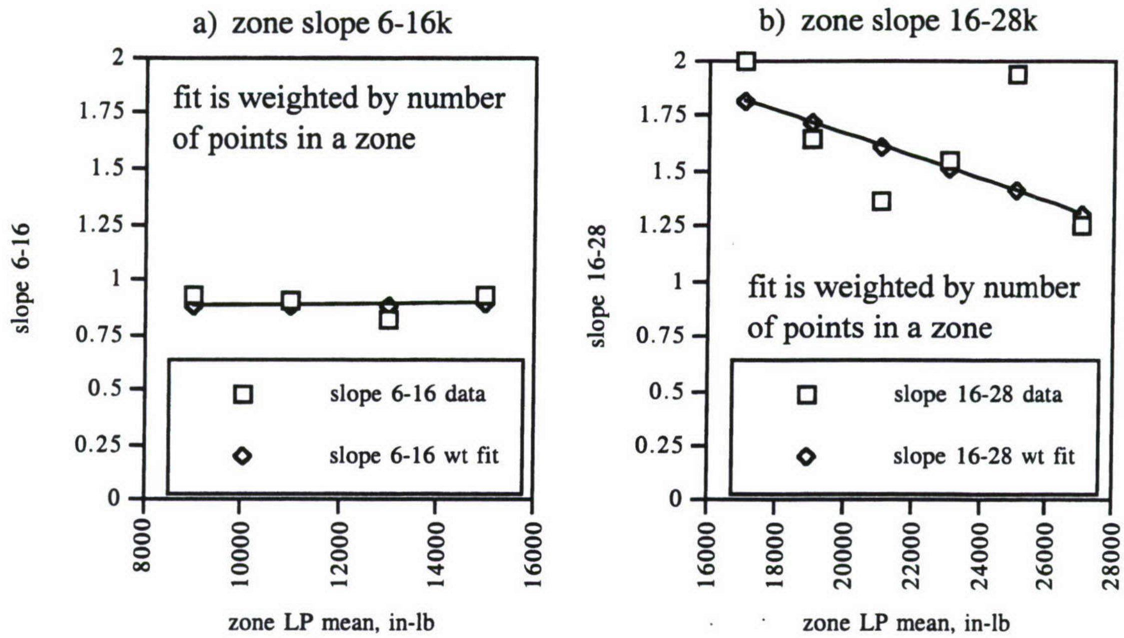


Figure 7-5. Vb Hp Fit for Slope by 2k LP Zones

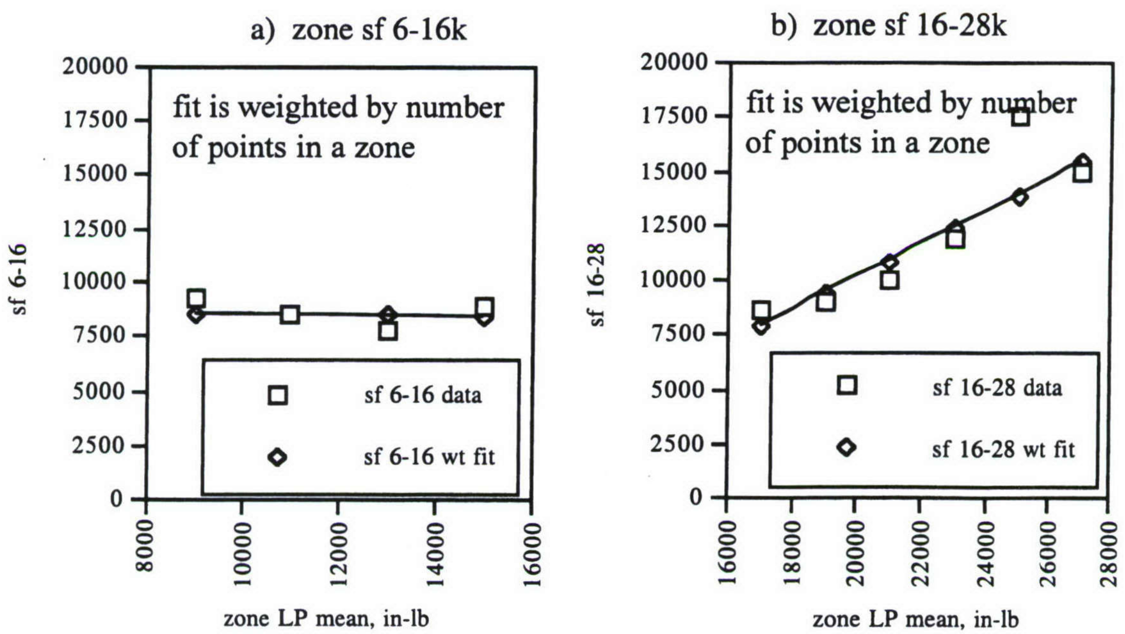


Figure 7-6. Vb Hp Fit for Scale Factor by 2k LP Zones

The fits shown are estimated by weighting each slope or scale factor zone value by the number of points in the zone as shown in Table 7-1.

Table 7-1. Number of Points in Each 2k Zone

LP Region 1: 6k - 16k		LP Region 2: 16k - 28k	
zone	number in zone	zone	number in zone
8k -10k	5	16k - 18k	21
10k - 12k	13	18k - 20k	19
12k - 14k	19	20k - 22k	24
14k - 16k	21	22k - 24k	17
		24k - 26k	6
		26k - 28k	5
total used	58	total used	92

Attention is directed to the distinct difference in character of the Weibull fits for each region.

In the LP 6k-16k region, both slope and scale factor have essentially constant values (Figures 7-5a and 7-6a) . The difference in the observed maximum value between 2k zones is due to the different number of samples in each 2k zone. The slope of 0.9 is close to the value of 1.0 of the exponential distribution often associated with slamming.

In the LP 16k-28k region, both slope and scale factor change with increasing LP value (Figures 7-5b and 7-6b). The slope decreases from approximately 1.8 to 1.2, while the scale factor increases by about a factor of 2. This suggests that not only is the phenomena changing from Rayleigh-like to exponential-like (becoming more slam-like), but that the severity is also increasing.

7.2.2 Ship B1 Sag Fitting Algorithm

Both low-pass and high-pass sag components are adequately fitted using a Weibull distribution:

$$P = 1 - \exp\left\{-\left[\left((x - x_0)/sf\right)^c\right]\right\} \qquad (1\text{-}1)$$

where

x_0	truncation or threshold value below which no events occur
sf	scale factor
c	slope: = 1, exponential; = 2, Rayleigh
x	variable of interest

7.2.2.1 Low-pass

The best fit Weibull parameters given in section 7.2.1.1 were modified by forcing X0LP = 4000 and then solving for the best fit parameter values for SFLP and CLP. The values used in combdist2[50] are: X0LP = 4000, scale factor SFLP = 15307.27928,[51] and slope CLP = 3.135234388.

7.2.2.2 High-pass

Region 1, low-pass < 16000 in-lb.

Weibull distribution with parameter X0BL = 0. SFL1 and CL1 are given by SFL1 = 8636.8229 - 0.012096286 * LP and CL1 = 0.860837149 + 1.90557×10^{-6} * LP where LP is the magnitude of the low-pass component.

Region 2, low-pass > 16000 in-lb.

Weibull distribution with parameter X0BH = 0. SFL2 and CL2 are given by SFL2 = -5006.969456 + 0.755259589 * LP and CL2 = 2.687986823 - 5.12231×10^{-5} * LP where LP is the magnitude of the low-pass component. If CL2 becomes less than 0.86, CL2 is set to 0.86.

Results are shown in Figure 8-1 (section 8.2).

7.3 Ship C

The author developed a method for estimating lifetime combined vertical and lateral loads (section 6.1) which makes use of a polar plot representation for the combined vertical and lateral loads so that the combined loads may be analyzed by polar plot sectors. This method is applied to Ship C ***full-scale*** data in ***ft-lt*** units scaled from model data.

As stated in section 7.1, "... , if a distribution fits, we are finished." We will stop analysis at the highest level at which we can determine a distribution.

We will start with a polar plot of the data, and see if we get a satisfactory fit for all the data in some relatively large sector (Figure 7-8). This works for "all combined magnitudes hog", but not for "all combined magnitudes sag".

We next select a sector (22.5 deg in this case) and try again. Again, this works for hog 22.5 magnitude, but not for sag 22.5 magnitude (Figure 7-10). We are now finished with hog (the Weibull space fit for hog changed very little from "all hog" to "hog 22.5"; compare Figure 7-8a with Figure 7-10a).

Next, resolve "sag 22.5 magnitude" into high-pass and low-pass components. When this is complete, it is seen that "sag 22.5 low-pass" has two different distributions (Figure 7-10b).

The final step is to find a distribution and fit for the "sag 22.5 high-pass component" as a function of "sag 22.5 low-pass component" (Figure 7-17).

---

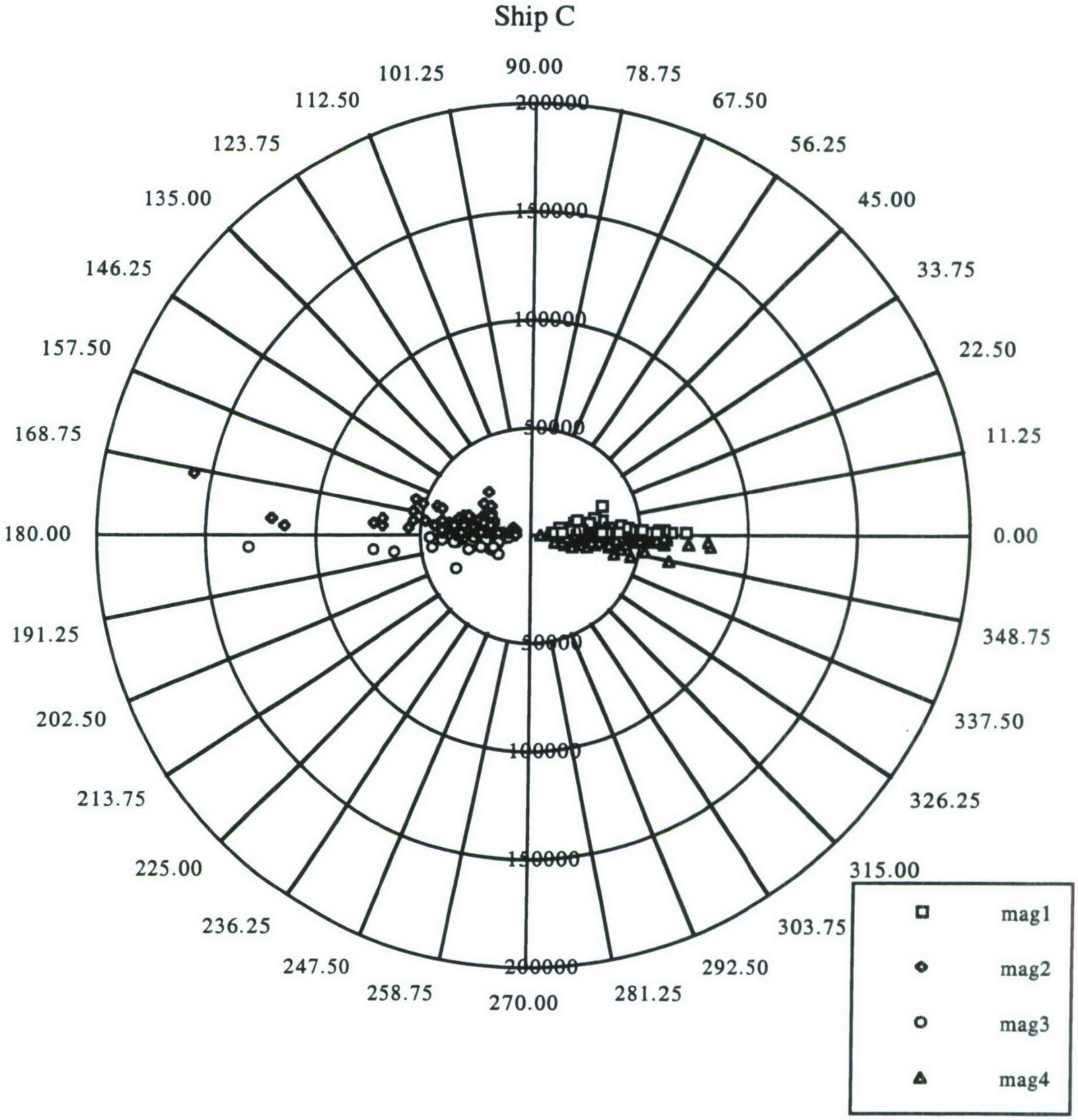
[50] "combdist2" is a computer program for evaluating combined distributions having combined low and high pass components. See section 8.1.1.

[51] More than the number of significant digits is given so that other people may check output values if they choose to write their own routines.

7.3.1 Determination of Highest Analysis Level For Fitting

<u>7.3.1.1 All Combined Magnitudes</u>

Figure 7-7 is a polar plot of combined vertical and lateral bending. The labels mag1, mag2, mag3 and mag4 refer to magnitudes in quadrants 1, 2, 3 and 4. The largest value of hog and sag was read in each independent interval (section 6.2) following a sign reversal from sag to hog or hog to sag. The times between these reversals correspond fairly well to the autocorrelation interval where the vertical bending autocorrelation function drops to about 0.1 of its peak value.



(Combined bending moments are in *ft-lt*; angles are in *degrees*)
(pure hog = 0 degrees, pure sag = 180 degrees)

Figure 7-7. Ship C: Polar Representation of Combined Vertical and Lateral Bending

The combined vertical and lateral hog magnitudes are about 1/2 of the combined vertical and lateral sag magnitudes.

Attention is called to the four largest sag points since these points are widely separated in magnitude from the remainder of the points. A wavelet analysis of the combined run strongly suggests that these points occur very close to the beginning or end of an individual run (information as to run segment lengths was not available). Examination of the run log showed no comments indicating that unusual behavior was noted during these individual runs which were then linked together to produce the combined run used for analysis.

The ln normal parameters of these four points plot fair with respect to the ln normal parameters of the other zones (Figures 7-17 and 7-18), and so it was decided to retain these points in the analysis. The retention and use of these points should be regarded with caution.

Weibull plots of all combined hog and sag magnitudes are shown below.

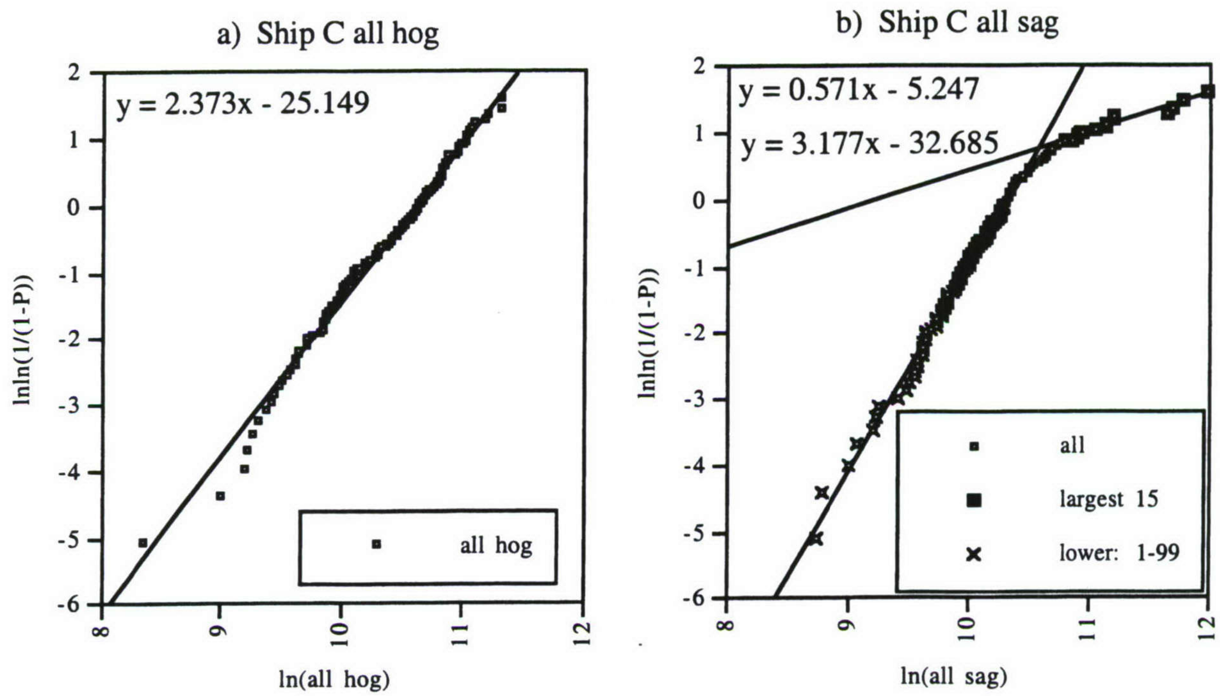


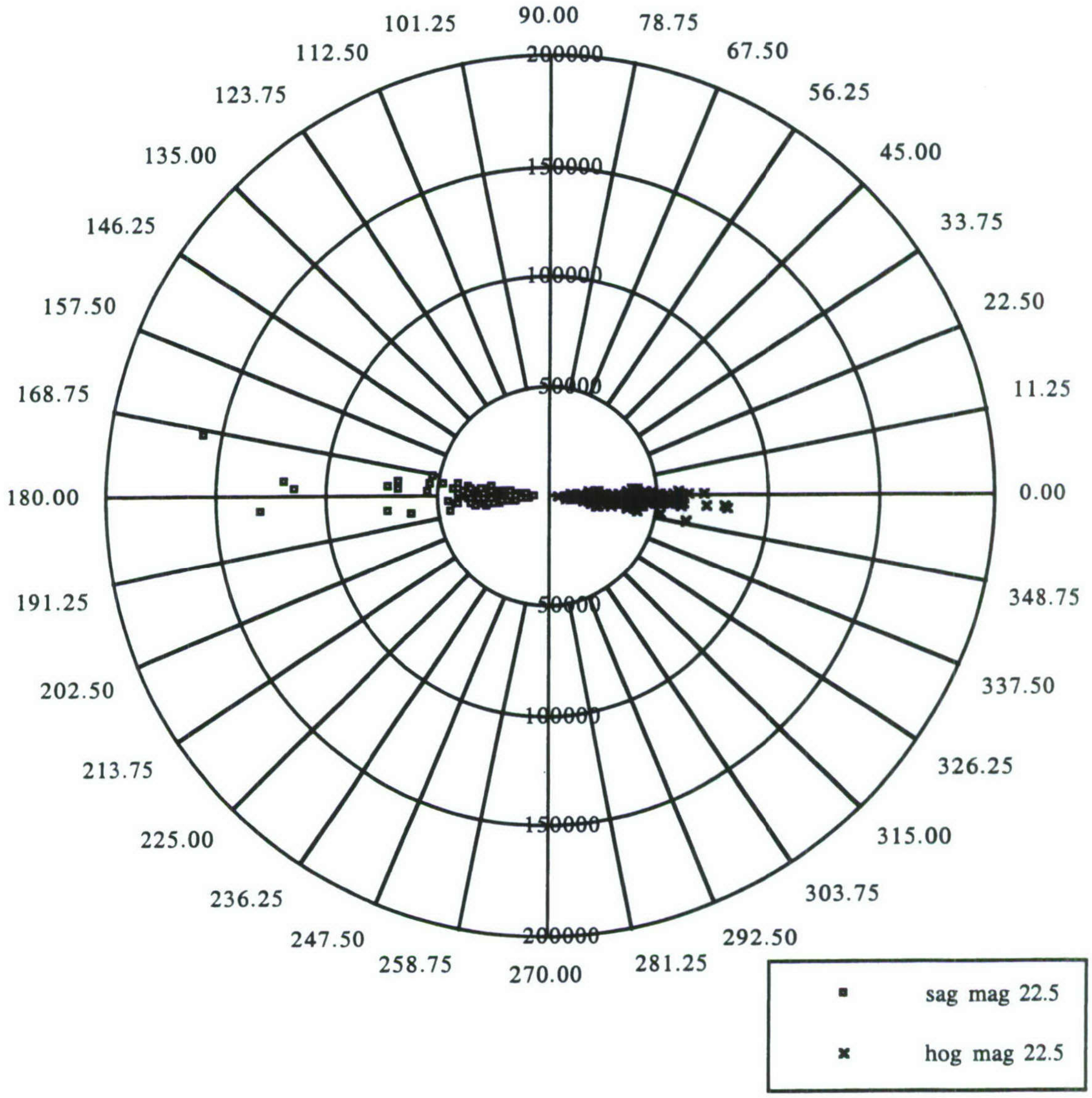
Figure 7-8. Ship C All Combined Hog and All Combined Sag Magnitudes

There will be no problem fitting hog: a single Weibull distribution looks adequate. Sag is a different story. There are at least two sag distributions.

7.3.1.2 Combined 22.5 deg Magnitudes

A sector 11.25 deg on each side of 180 deg (180 deg is pure sag, i.e., no lateral bending component) and on each side of 0 deg (0 deg is pure hog), i.e., no lateral bending component) was selected for the main analysis. These points are shown in Figure 7-9.

Ship C: Hog, Sag Points in 22.5 deg Sectors



(Combined bending moments are in *ft-lt*; angles are in *degrees*)
(pure hog = 0 degrees, pure sag = 180 degrees)

Figure 7-9. Polar Representation of Combined Vertical and Lateral Bending, 22.5 Deg Sector

Weibull plots of the 22.5 deg sectors combined hog and sag magnitudes are shown in Figure 7-10. In Figure 7-10b, the upper equation is the result of fitting the largest 18 combined sag points in Weibull space, the lower equation is the result of fitting the smallest 85 combined sag points in Weibull space.

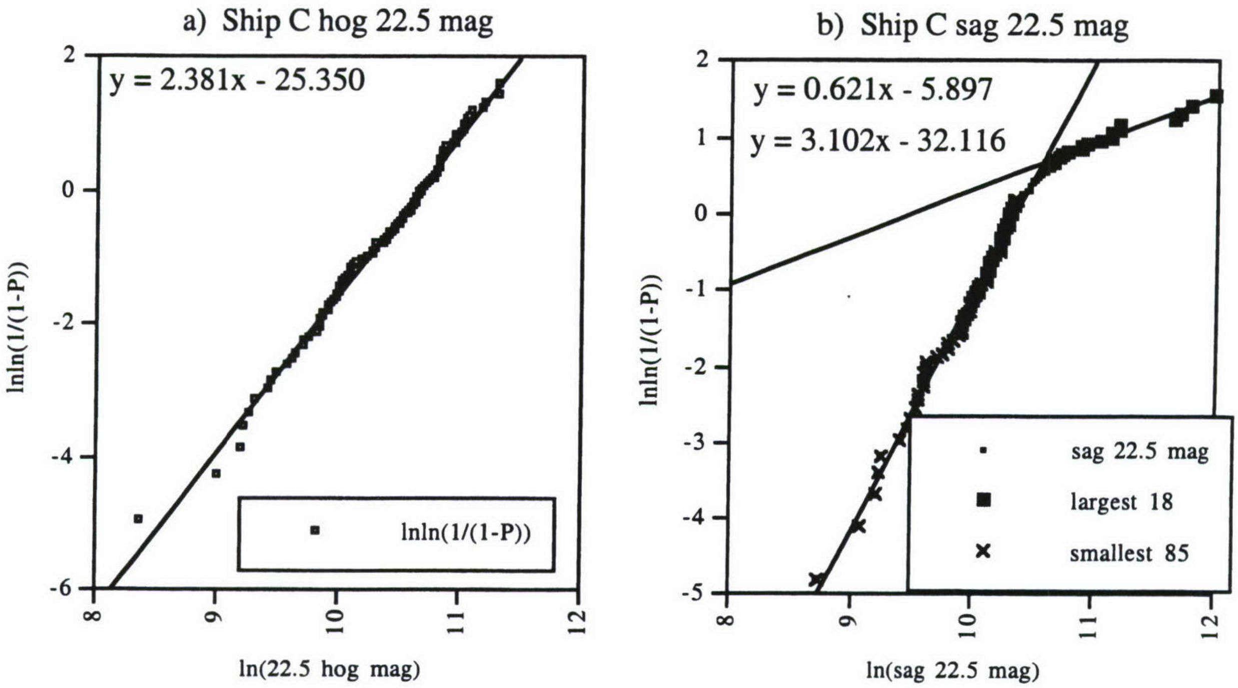


Figure 7-10. Combined Hog and Combined Sag, 22.5 Sector

Again, there will be no problem fitting hog: a single Weibull distribution looks adequate. Sag is still a different story. There are at least two sag distributions.

7.3.1.3 Resolve Sag 22.5 deg into Components

The next step is to find the vertical bending sag magnitude and then resolve these into low-pass and high-pass components. The result is shown in Figure 7-11. In Figure 7-11a, the upper equation is the result of fitting the largest 6 vb lp sag points in Weibull space, the lower equation is the result of fitting the smallest 114 vb lp sag points in Weibull space.

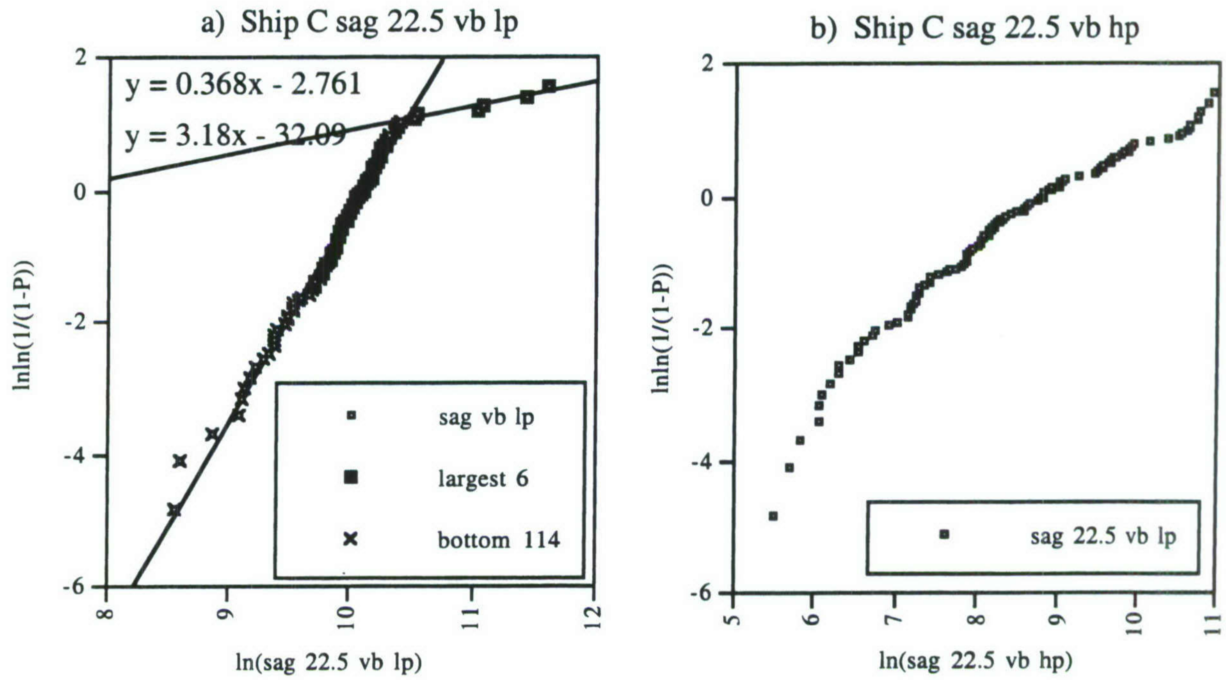


Figure 7-11. Sag 22.5 Vb Low-pass and Vb High-pass

There appear to be two vertical bending low-pass sag distributions. The vertical bending high-pass sag distribution (Figure 7-11b) is not Weibull (note the curve in Weibull space - if a distribution is Weibull, it plots as a straight line in Weibull space).

7.3.2 Determining Fits for Sag 22.5 Vertical Bending Components

7.3.2.1 Determining Fits for Sag 22.5 Vertical Bending Low-pass

There are two domains for vertical bending low-pass: the smallest 114 points and the largest 6 points.

Bottom Set of 114

Figure 7-12 shows the smallest 114 low-pass points.

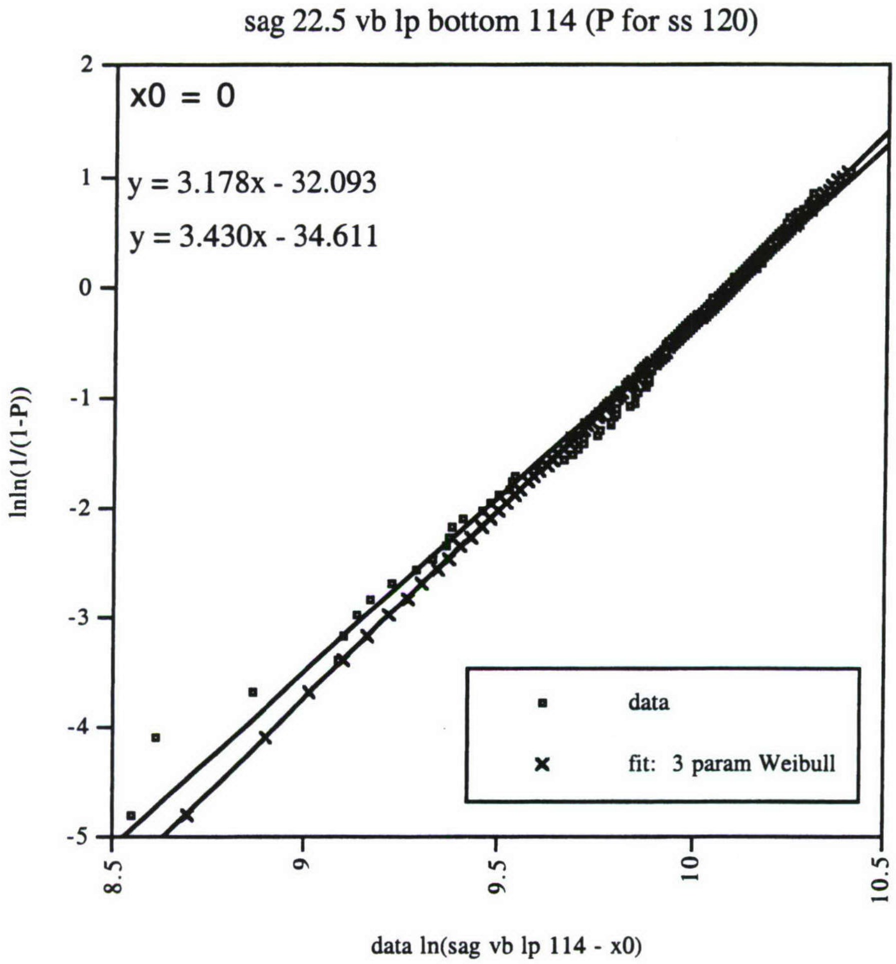


Figure 7-12. Weibull Plot 114 Smallest Sag Vb Lp

The 3 parameter Weibull fit parameters (which will be used in combdist2) were estimated in original space. The ln of the smallest points distorted the slope (3.43 original space fit, 3.18 Weibull space). The sample size of 114 is plotted as coming from the original sample size of 120; an argument could be made that it should be plotted as a sample of 114. When used in combdist2 its pdf will be multiplied by 0.942 which is the ratio 114 / (120+1).

An argument could be made that the multiplier should be 114 / 120 = 0.95.

Top Set of 6

Figure 7-13 shows the 6 largest vb lp sag points.

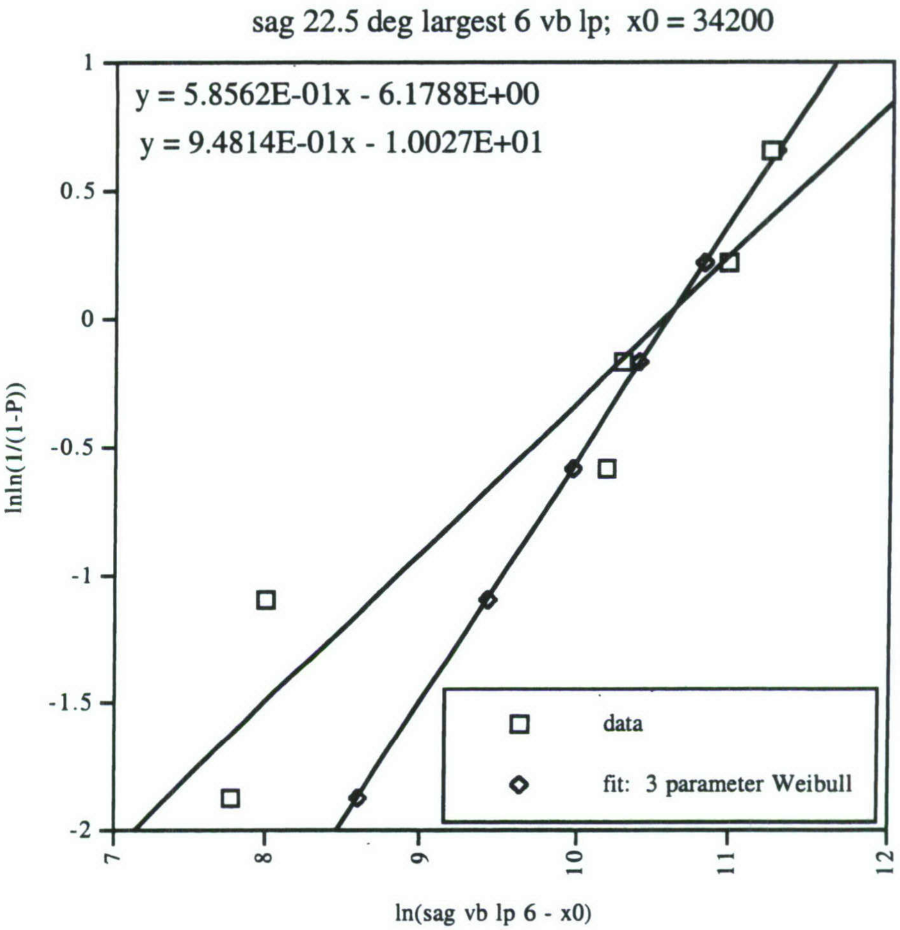


Figure 7-13. Weibull Plot 6 Largest Sag Vb Lp

The 3 parameter Weibull fit parameters (which will be used in combdist2) were estimated in original space. The ln of the smallest points significantly distorted the slope (0.948 original space fit, 0.586 Weibull space). This sample size of 6 is plotted as coming from a sample size of 6. When used in combdist2, its pdf will be multiplied by 1 - 0.942 = 0.058. If the sample size is taken to be 120, the multiplier will be 0.05.

The x_0 value 34200 ft-lt is the intersection value of the two vb lp lines in Figure 7-11a.

This is getting to be a very small sample size. The Weibull parameters should be re-estimated using weights on the points.

7.3.2.2 Determining Fits for Sag 22.5 Vertical Bending High-pass

Figure 7-14 is a plot of vertical bending low-pass vs. high-pass. The vertical lines are at 5000 ft-lt intervals.

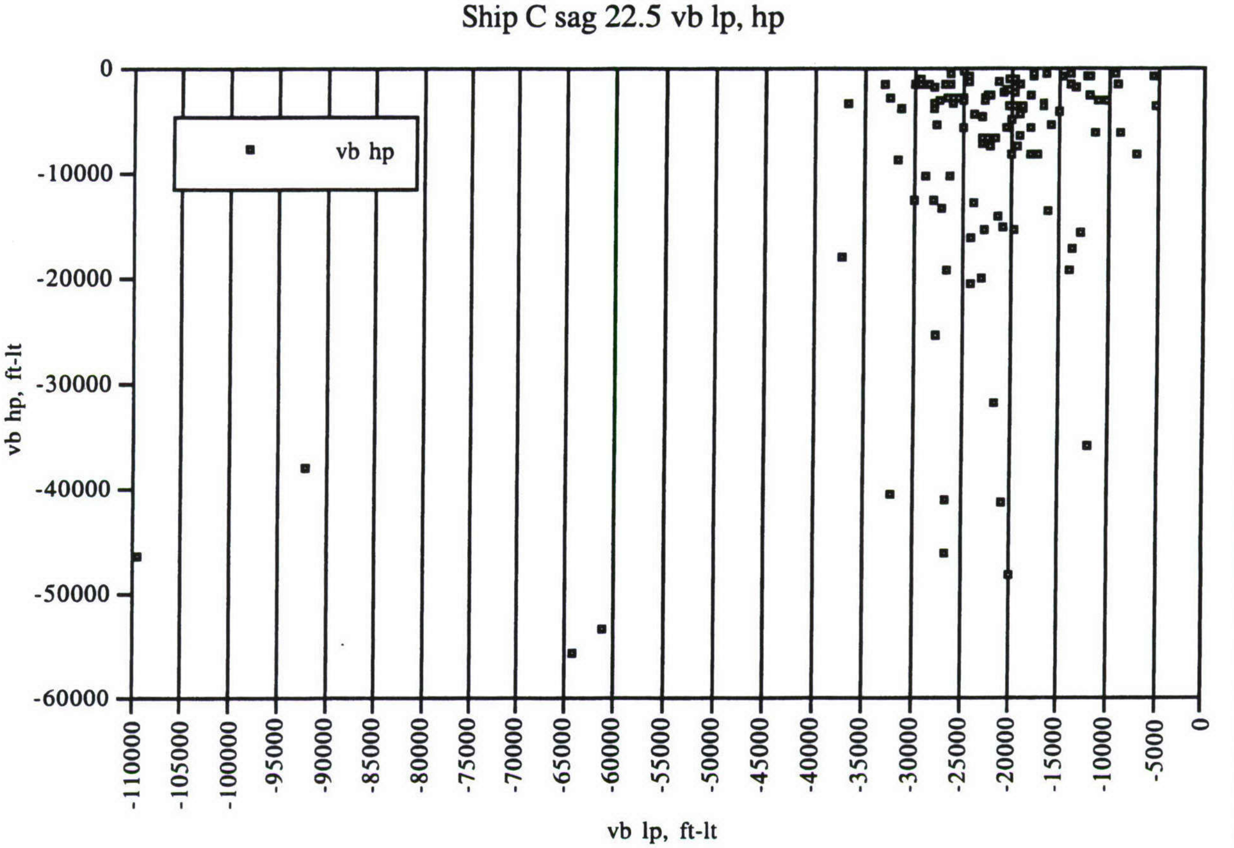


Figure 7-14. Vertical Bending Low-pass vs. High-pass

We will now attempt to find distributions for each approximately 5000 ft-lt wide zone. The number of points in each zone is shown in Table 7-2.

Table 7-2. Lp Zone Centers Using Events

lp zone ft-lt	number of hp events	weighted lp zone center	unweighted lp zone center
5k-10k	7	7450	7500
10k-15k	15	12612	12500
15k-20k	25	18656	17500
20k-25k	35	22209	22500
25k-30k	27	27195	27500
30k-40k	7	33342	35000
60k-110k	4	79648	85000

7.3.3 Weibull Fits

The first recourse is to assume Weibull distributions. Figure 7-15 shows Weibull plots by lp zones below and on the next page.

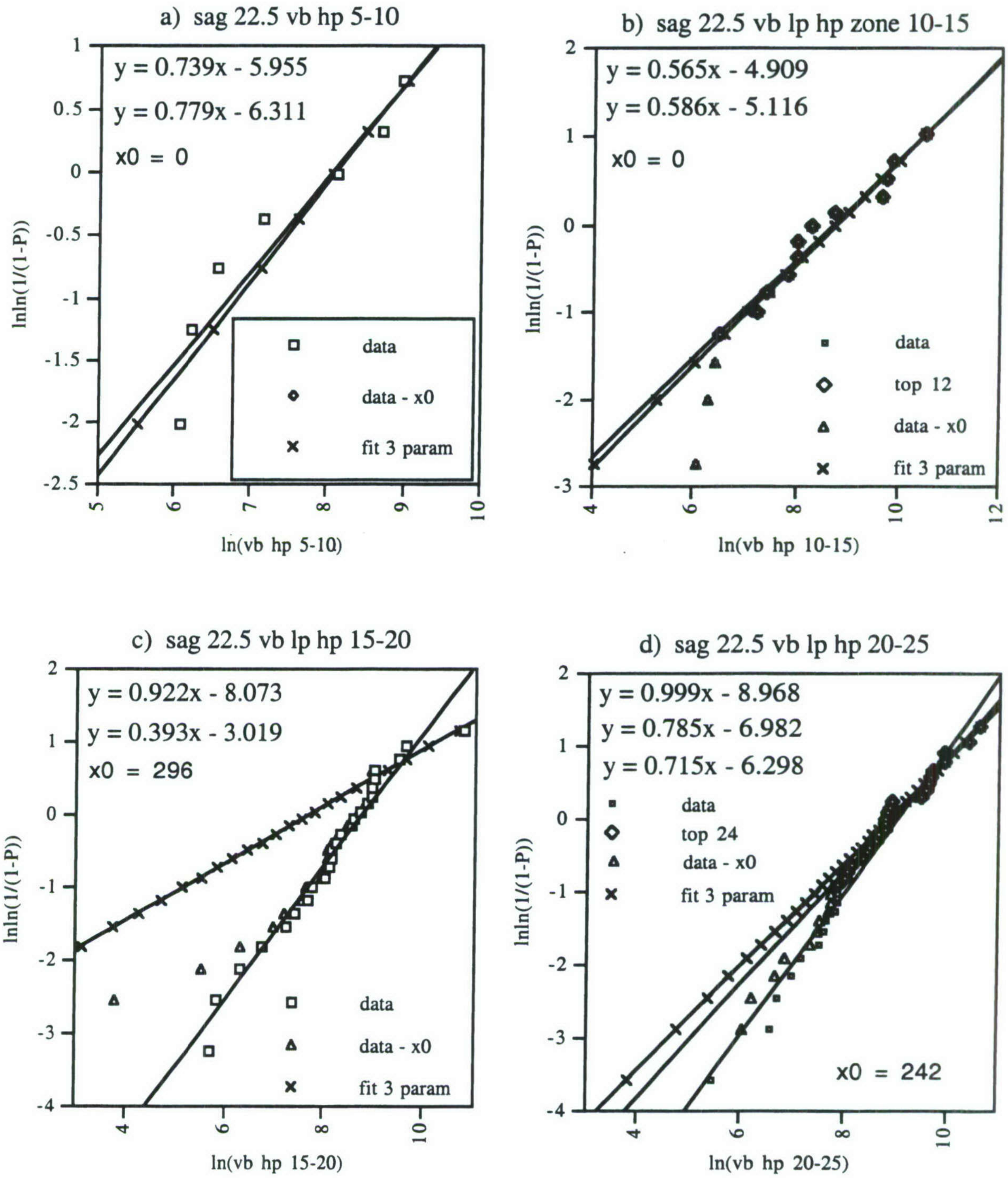


Figure 7-15. Sag 22.5 Vb Hp by Lp Zones

Note that for vb hp zone 15-20 the largest four points drive the 3 parameter fit.

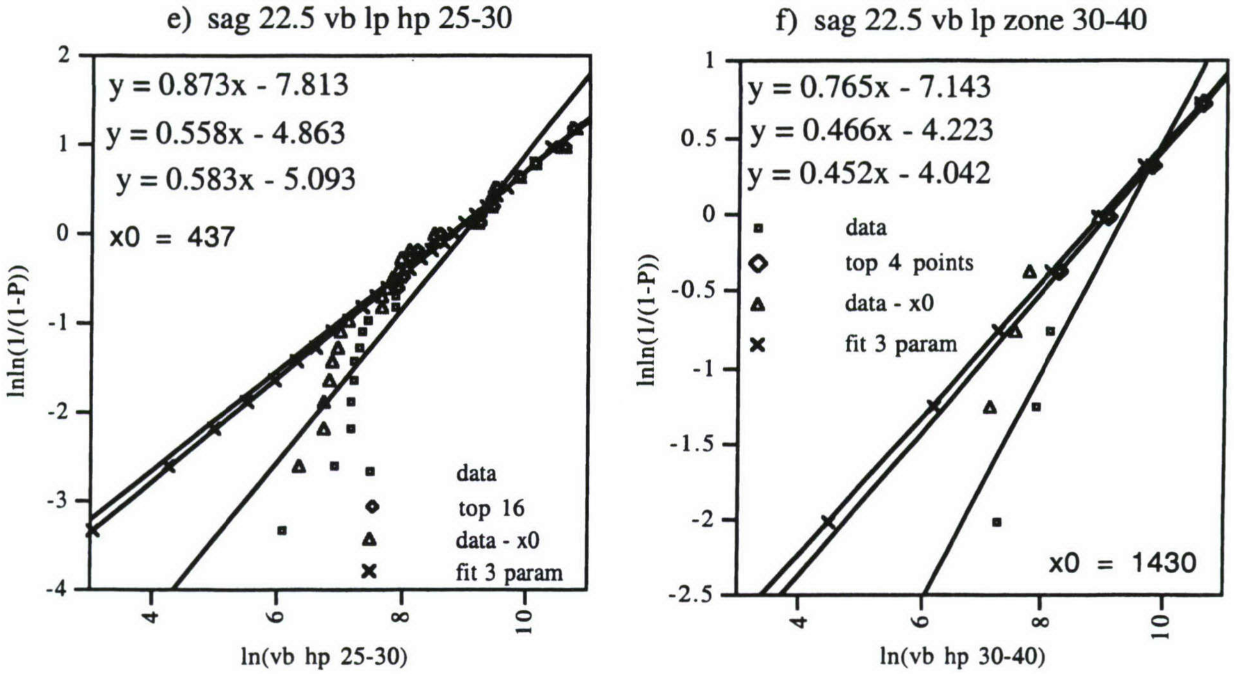


Figure 7-15. Sag 22.5 Vb Hp by Lp Zones (continued)

If the data is Weibull, it will plot as a straight line on a Weibull plot. There is a general tendency for the data to curve, and to curve in a direction which is non-conservative if a single line were fitted using all the points. Further, some of zones show more than one slope.

The 3 parameter Weibull fits show a strong tendency to fit the upper slope if there are two distributions.[52] Further, these upper slopes are unrealistically low - an exponential distribution has a slope = 1. Slopes below one have been observed, but they are usually not below 0.8.

The conclusion is the Ship C hp data cannot be satisfactorily fitted using Weibull distributions. We now try another distribution type.

[52] This is reasonable behavior since in original space the least squares objective function is more strongly driven by large differences between the data and the fit. This is more likely to happen where the data values are large. Because we are using order statistics, these large data values occur at the upper end.

7.3.4 Log Normal Fits

The log normal distribution has pdf given by

$$\mathrm{pdf}(x) = \frac{1}{\sqrt{2\pi} * \varsigma * x} e^{\left[-\frac{1}{2}\left(\frac{\ln x - \lambda}{\varsigma}\right)^2\right]} \qquad 0 \le x < \infty \tag{7-1}$$

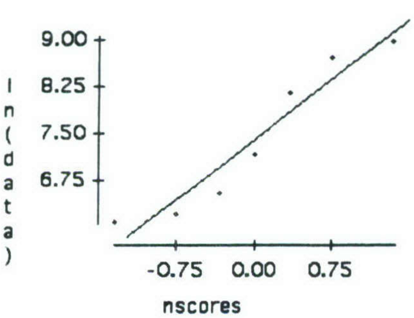
where

x	original data (vb hp component value)
λ	mean of the ln(x) values
ς	standard deviation of the ln(x) values

Log normal plots and statistics are shown in Figure 7-16. In these figures the mean is λ and the standard deviation is ς. For a perfect fit to the log normal distribution, both the skewness and kurtosis = 0 for the distribution of the natural logarithm of the data..

Some of the figures on the following pages have two columns. The second column has the log normal fit for the 3 parameter Weibull fit. Where there is no second column, $x_0 = 0$ and the original data and the $x\text{-}x_0$ data are the same, and so the fit is the same.

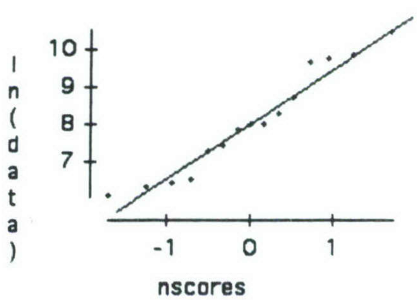
a) sag 22.5 vb hp 5-10 x0=0



Summary of **ln(data)**
No Selector

Total Cases	7
Count	7
Mean	7.40964
Median	7.16470
Variance	1.46371
StdDev	1.20984
Range	2.91780
Skewness	0.186889
Kurtosis	-1.61424
StdErr	0.457277

b) sag 22.5 vb hp 10-15 x0=0



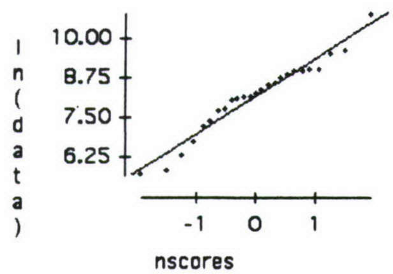
Summary of **ln(data)**
No Selector

Total Cases	15
Count	15
Mean	8.04247
Median	8.00240
Variance	1.99752
StdDev	1.41334
Range	4.38940
Skewness	0.234814
Kurtosis	-1.12903
StdErr	0.364922

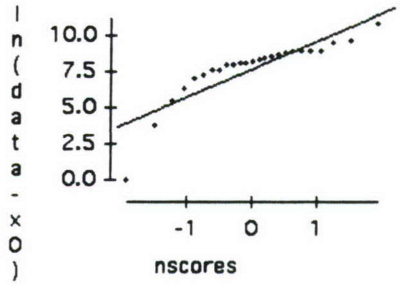
Figure 7-16. Ln Normal Fit Hp for Lp Zones

c) sag 22.5 vb hp 15-20

x0 = 0



x0 = 296



Summary of **ln(data)**
No Selector

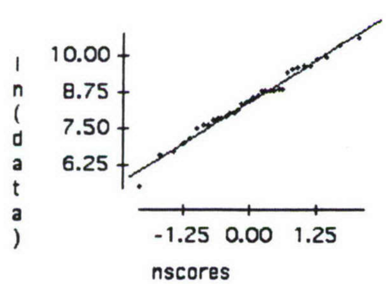
Total Cases	25
Count	25
Mean	8.18134
Median	8.24880
Variance	1.40277
StdDev	1.18439
Range	5.08750
Skewness	-0.348722
Kurtosis	0.136567
StdErr	0.236877

Summary of **ln(data-x0)**
No Selector

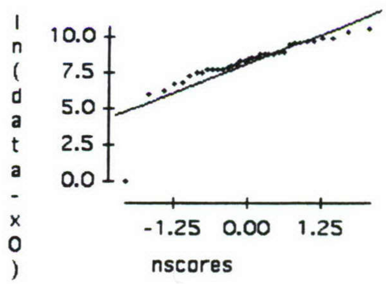
Total Cases	25
Count	25
Mean	7.75888
Median	8.16820
Variance	4.54756
StdDev	2.13250
Range	10.7750
Skewness	-2.16820
Kurtosis	5.42204
StdErr	0.426500

d) sag 22.5 vb hp 20-25

x0 = 0



x0 = 242



Summary of **ln(data)**
No Selector

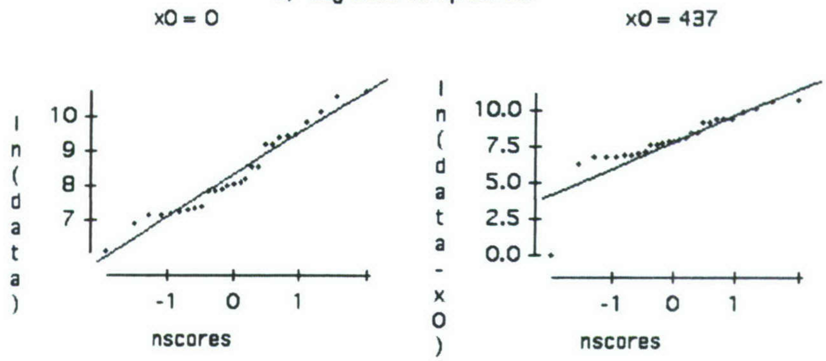
Total Cases	35
Count	35
Mean	8.43573
Median	8.45380
Variance	1.27427
StdDev	1.12884
Range	5.13440
Skewness	-0.266068
Kurtosis	-0.009528
StdErr	0.190808

Summary of **ln(data-x0)**
No Selector

Total Cases	35
Count	35
Mean	8.17665
Median	8.38870
Variance	3.24871
StdDev	1.80242
Range	10.6203
Skewness	-2.65225
Kurtosis	10.2185
StdErr	0.304664

Figure 7-16. Ln Normal Fit Hp for Lp Zones (continued)

e) sag 22.5 vb hp 25-30



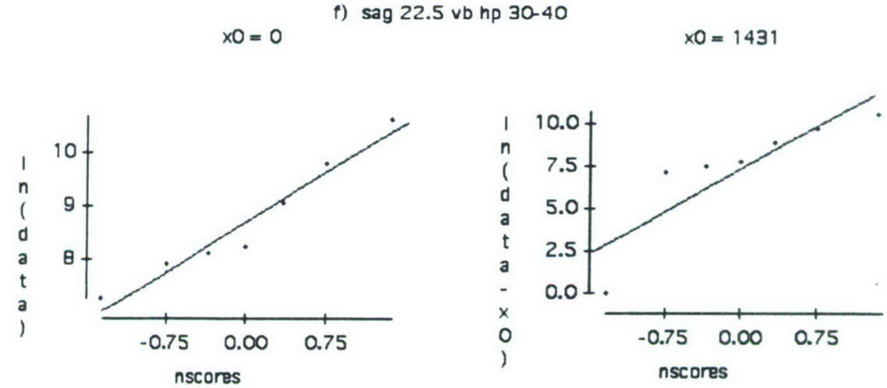
Summary of ln(data)
No Selector

Total Cases	27
Count	27
Mean	8.33486
Median	8.05390
Variance	1.45913
StdDev	1.20794
Range	4.65580
Skewness	0.382591
Kurtosis	-0.717384
StdErr	0.232469

Summary of ln(data-x0)
No Selector

Total Cases	27
Count	27
Mean	7.93571
Median	7.90430
Variance	4.10666
StdDev	2.02649
Range	10.7285
Skewness	-2.07335
Kurtosis	6.82235
StdErr	0.389998

f) sag 22.5 vb hp 30-40



Summary of ln(data)
No Selector

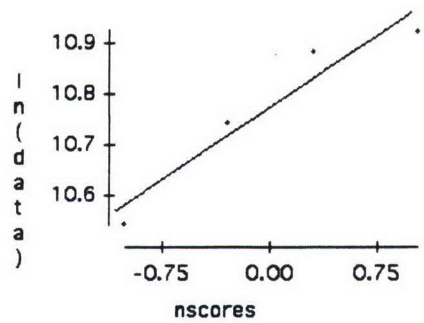
Total Cases	7
Count	7
Mean	8.70903
Median	8.24280
Variance	1.36647
StdDev	1.16896
Range	3.33710
Skewness	0.478183
Kurtosis	-1.00051
StdErr	0.441826

Summary of ln(data-x0)
No Selector

Total Cases	7
Count	7
Mean	7.37149
Median	7.77060
Variance	12.0842
StdDev	3.47623
Range	10.5671
Skewness	-1.50736
Kurtosis	1.11699
StdErr	1.31389

Figure 7-16. Ln Normal Fit Hp for Lp Zones (continued)

g) sag 22.5 vb hp 60-100 x0 = 0



Summary of **ln(data)**
No Selector

Total Cases	4
Count	4
Mean	10.7733
Median	10.8129
Variance	0.029543
StdDev	0.171881
Range	0.381200
Skewness	-0.566056
Kurtosis	-1.23324
StdErr	0.085941

Figure 7-16. Ln Normal Fit Hp for Lp Zones (continued)

The logs of the original data ($x_0 = 0$) plot better as log normal than do the logs of the 3 parameter Weibull fits (x_0 ne 0).

The log normal fit of the original data has the additional advantage that there is no threshold or truncation value.

The log normal fits are superior to either the Weibull fits of the original data or the Weibull 3 parameter fits, and will be used for fitting the Ship C vertical high-pass data.

The next step is to form equations for the variation with vertical bending low-pass component for the mean and the standard deviation of the high-pass component. Plots of the mean (Figure 7-17) and the standard deviation (Figure 7-18) are shown along with the associated linear fit.

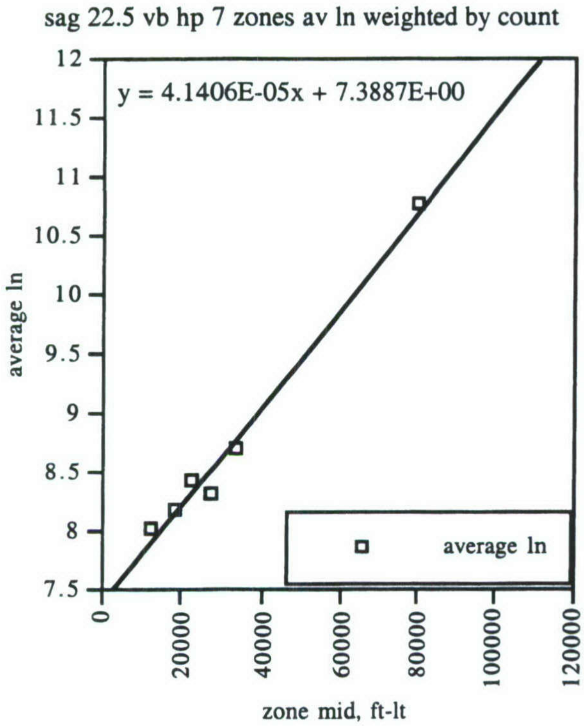


Figure 7-17. Avg Ln Weighted by Counts

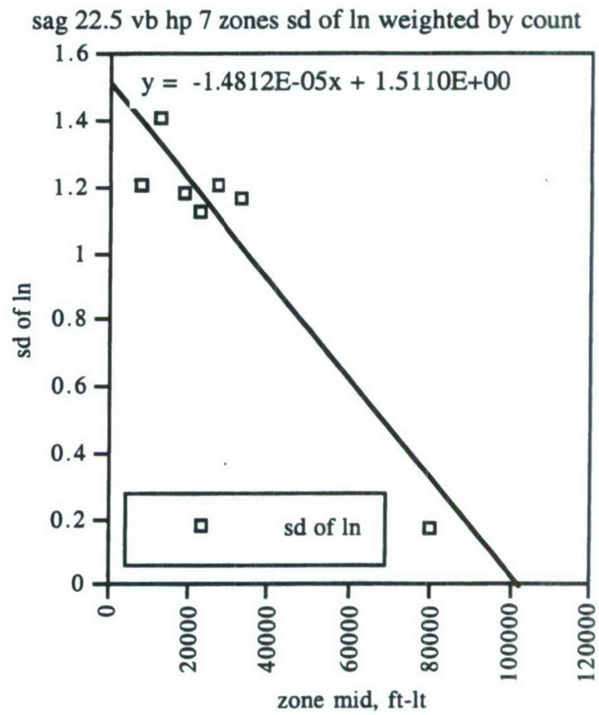


Figure 7-18. St Dev Ln Weighted by Counts

A ln mean or ln standard deviation fit is weighted by the number of events in its zone when finding the equation of the straight line.

The ln standard deviation fit needs to be truncated before it reaches zero. A minimum value of 0.1 was selected as the minimum value of the ln standard deviation.

7.3.5 Ship C Sag Fitting Algorithm

7.3.5.1 Low-pass

The low-pass component is fitted using two Weibull distributions.

Lp Value < 33,244 ft-lt

The values used in combdist2 are: X0LP1 = 0, scale factor SFLP1 = 24123.10395, and slope CLP1 = 3.429892869.

The low-pass pdf for low-pass values < 33,244 is multiplied by 0.942.

Lp Value > 33,244 ft-lt

The values used in combdist2 are: X0LP2 = 33,244, scale factor SFLP2 = 40280.9993, and slope CLP2 = 0.971556831.

The low-pass pdf for low-pass values > 33,244 is multiplied by 0.058.

7.3.5.2 High-pass

Log normal distribution with parameters AV and SD.

Average

AV = AVB1 + AVM1 * X

where

AVB1	7.3887D+00
AVM1	4.1406D-05
X	vertical bending low-pass magnitude

Standard Deviation

SD = SDB1 + SDM1 * X

where

SDB1	1.5110D+00
SDM1	-1.4812D-05
X	vertical bending low-pass magnitude

If the computed value of the standard deviation is less than 0.1, the value of the standard deviation is set = 0.1.

Results are shown in Figure 8-2 (section 8.2).

8.0 COMBINED FITS

Graphical examples showing goodness of fit for the Ship B and Ship C model test runs analyzed in section 7.2 (Ship B), and section 7.3 (Ship C) are shown in section 8.2. Comments on the computer program developed to evaluate the combined distributions are provided in section 8.1.

8.1 Computational Considerations

8.1.1 Combdist2

A computer program, combdist2, was developed to apply the method(s) of combining low-pass and high-pass components to form a combined load. It can be operated in two modes:

1) computational mode: numerical integration is used to compute the probability density function (pdf) of the combined low-pass and high-pass components, or

2) simulation mode: simulation is used to estimate the probability density function (pdf) of the combined low-pass and high-pass components.

Combdist2 is organized to facilitate the addition of other ships, tests, test conditions, and runs as they become available.

8.1.2 Computational Mode

Combdist2 evaluates a double integral. The inner integral finds pdf(t) associated with a constant value of $t = lp + hp$. The outer integral integrates pdf(t) over the values of t to find P.

In the course of performing these numerical integrations, improper integrals were sometimes encountered. These can be of two types for our problem:[53] a) the value of the function goes to ∞ at a finite limit of integration, or b) the value of the function goes to zero at a limit of integration which is infinite. In either case, the value of the integral must stay finite (or else the value of the integral is infinite). Techniques were developed to deal with both of these types of impropriety.

8.1.2.1 Type A Impropriety

The first type (infinite function at a finite boundary) occurred for the Ship B example when evaluating the inner integral [find pdf(t)] for $t = x + y$). In this case it is due to a Weibull distribution with a slope less than one [footnote 31 (section 5.0)].

The usual fixes cannot be applied since the kernels of the integrals are not amenable to analytic integration. Consequently, recourse is made to numerical methods.

[53] Improper integrals also occur when a singularity occurs in the range of integration. So far, this has not happened with the functions used in the examples.

The fix is: make use of the fact that for the value of the integral to stay finite, while the function goes to infinity at a limit, means that incremental areas go to zero faster than the function can go to infinity.

The technique is:

1) Pick a point near the nearest boundary at which the function goes to infinity, say at 0.999 of the integration range, so we are left with integrating over 0.001 of the range after integrating over the first 0.999 of the range.

2) Form a subregion = 1/2 of this 0.001 range starting at 0.9990 and ending at 0.9995. Find the value of the integral in this region and call it da(1).

3) Form a subregion = 1/4 of the 0.001 range starting at 0.9995 and ending at 0.99975. Find the value of the integral in this region and call it da(2).

4) Form a subregion = 1/8 of the 0.001 range starting at 0.99975 and ending at 0.999875. Find the value of the integral in this region and call it da(3).

5) Continue in this fashion until da(n) is less than some cutoff value.

6) If the sequence da = da(1) + da(2) + da(3) + ... + da(n) converges add da to the result of integrating over the first 0.999 of the range (if it doesn't, the integral may be impossible rather than improper).

8.1.2.2 Type B Impropriety

The second type (zero function at an infinite boundary) occurred in the Ship C example when integrating the outer integral (over all values of $t = x + y$). (In this case, it is due to a log normal distribution with a constantly increasing mean value.)

The usual fixes cannot be applied since the kernels of the integrals are not amenable to analytic integration. Consequently, recourse is made to numerical methods.

The fix is:

1) Find the pdf of $t = lp + hp$ to some large finite value.

2) Integrate pdf(t) to obtain some value Pnn.

3) Divide the pdf(t) values found in step 1 by (Pnn + some small value such as 10^{-15}) to form normalized pdf(t) values (by definition probability values lie between 0 and 1).

4) These values will be useful over some wide range of $t = lp + hp$ values. The limits of applicability may be estimated by Weibull plotting [ln(t) vs. lnln(1/(1-P))] the t values vs. the P values using the normalized pdf(t) values found in 3) above, and then finding the value of the uppermost inflection point. Use this value as an estimate of the largest $t = lp + hp$ value to use in further computations.

8.1.2.3 Normalization

Step 3) above is a normalization step used in both computation and simulation mode irrespective of whether improper integrals occurred. It is imperative for simulation mode.

With ideal function specification and integration techniques, the result of the integration before normalization should be very close to 1.0. Trapezoidal integration is used in computational mode since there may be variable $t = x + y$ spacing. In simulation mode, the current implementation uses constant increments of $t = x + y$.

In computation mode a check problem with constant t spacing results in a value before normalization of Pnn = 0.9999117, while with variable t spacing the result is 1.0017866. The computational mode results for Ship B and Ship C are not as good as those for the check problem. As will be seen in section 8.2, the results using normalization are good whether or not computational or simulation mode is used; so, while this lack of computational mode convergence to 1 is not yet explained, its effect on the work reported here does not seem to be significant.

8.1.3 Simulation Mode

The (pseudo) random number generator (rng) used is due to Knuth[54] and is available on his web site (http://www-cs-faculty.stanford.edu/~knuth/). Further remarks on this type of rng are available in [Knuth, 1998]. It has a very long period[55] and so the chances of repeating are essentially zero[56] for the type of problem we have.

Simulation mode won't eliminate either type of impropriety. However, the chances of encountering a type A (infinite function at a finite boundary) impropriety are very small. The same comments as above for type B (zero function at an infinite boundary) impropriety apply. The solution is to take a large range over which to do the simulation. While the same test as in section 8.1.2.2 step 4) may be used, the result is misleading since the randomness in the second derivative makes the first inflection point from above occur at too large a value of $t = lp + hp$.

Because we wish to see the effect of using different ranges, the same seed was used for all examples. The same number of random numbers (rn) [$50*10^6$ pairs of lp and hp] was used for all examples. This means that the density/(unit of range) varies[57]. For production, rather than illustration runs, both seeds and rn/range need to be varied.

---

54 The double precision version of his lagged Fibonacci generator is used. It has the remarkable property that it will give a unique sequence of rn when started with any integer between 0 and 2^{30}-3 = 1,073,741,821 inclusive. The suggestion that only the first 100 of each set of 1009 rn be used (throw away the remaining 909 rn) has been implemented in combdist2.

55 The period before repeating for the integer version is about $2.7*10^{39}$. The double precision version used here has a repeat period about $1.4*10^{45}$. Recall that a billion is 10^9, a trillion is 10^{12}.

56 Known to be true for sequence length $< 2^{70} \sim 1.18*10^{21}$.

57 A more consistent approach would be to use the same number of rn/(unit of range).

8.2 Combined Fit Examples

Figures 8-1 and 8-2 show sag combined vertical bending moment for Ship B and Ship C respectively.

8.2.1 Ship B

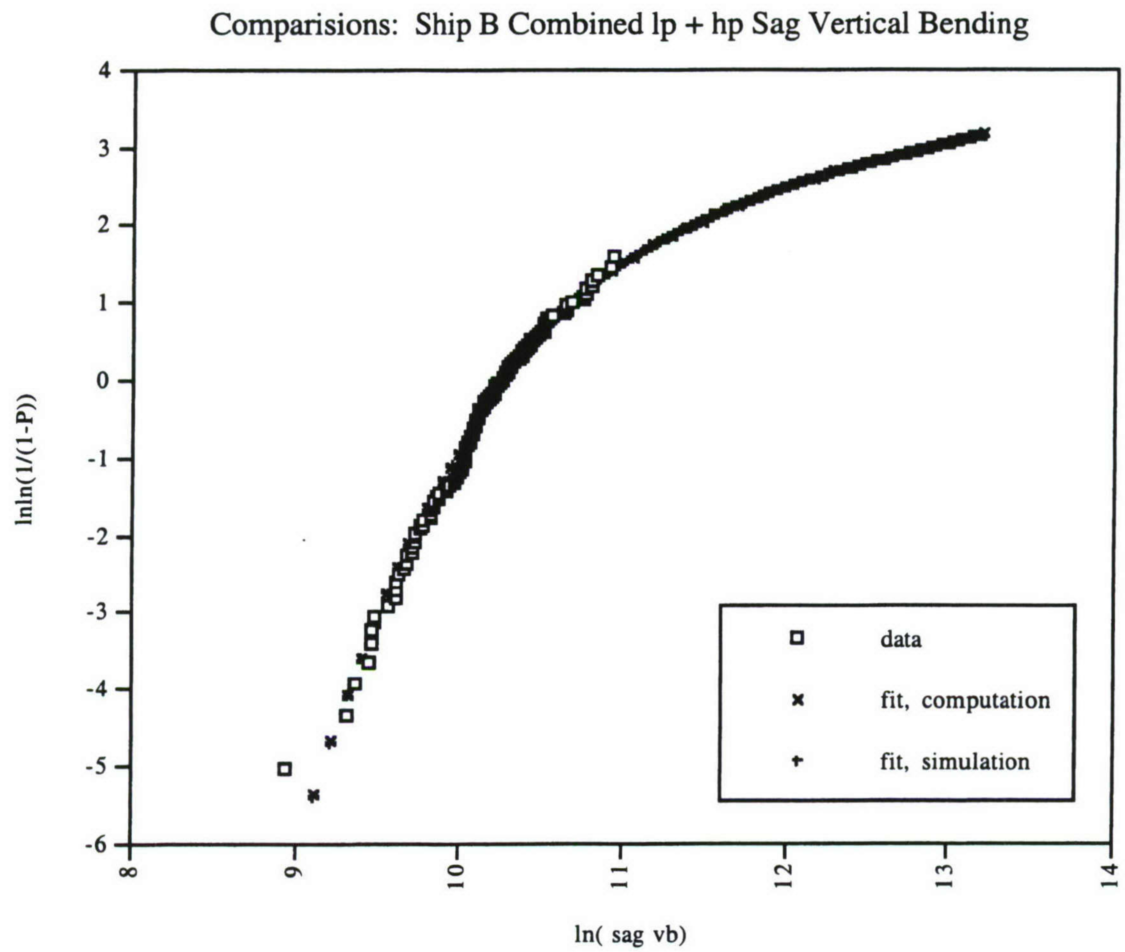


Figure 8-1. Ship B Sag Combined Vertical Bending

The computation and simulation results overlay each other and both fit the data well. The full range of the fit may be used for further computation such as that required for extreme values.

The initial bin width for computation was 1000 in-lb / bin. The automatic adjustment for bin size was used in computation mode.

8.2.2 Ship C

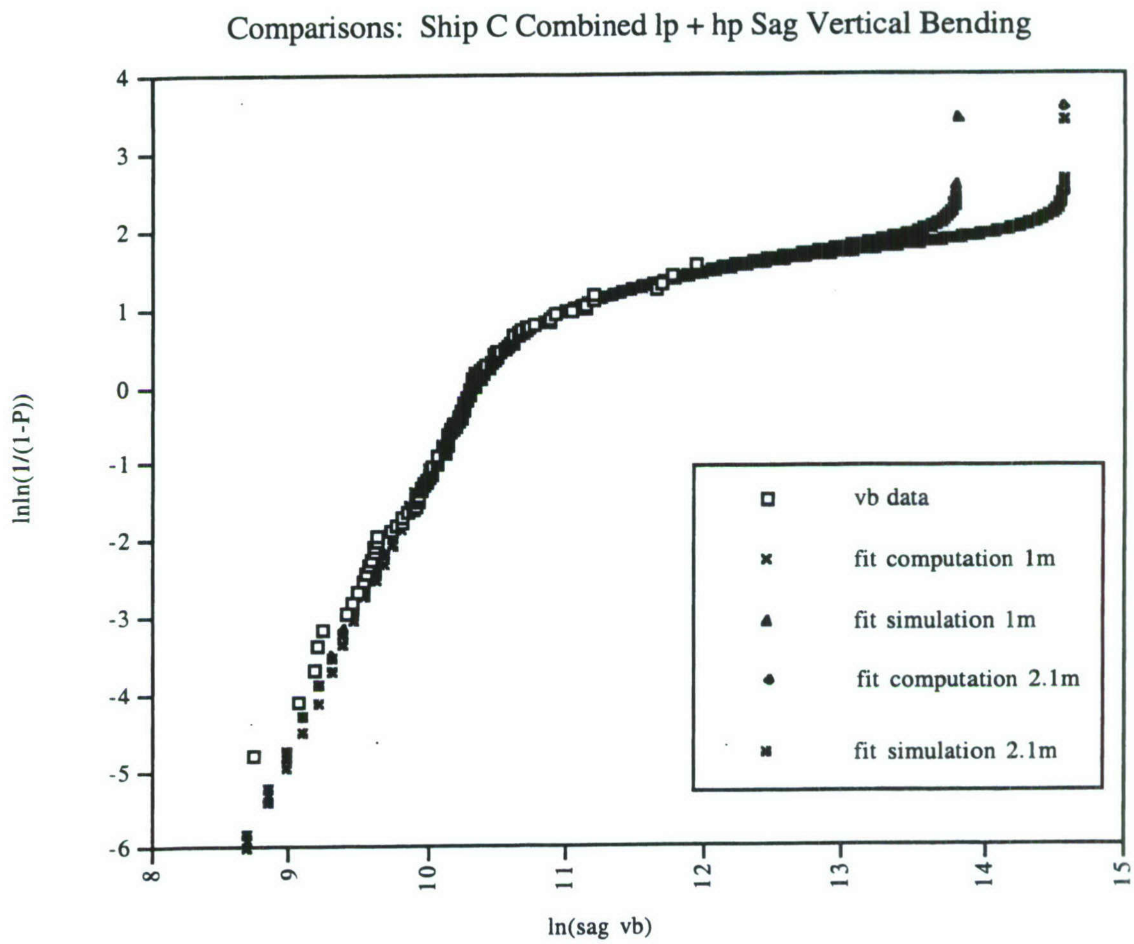


Figure 8-2. Ship C Sag Combined Vertical Bending

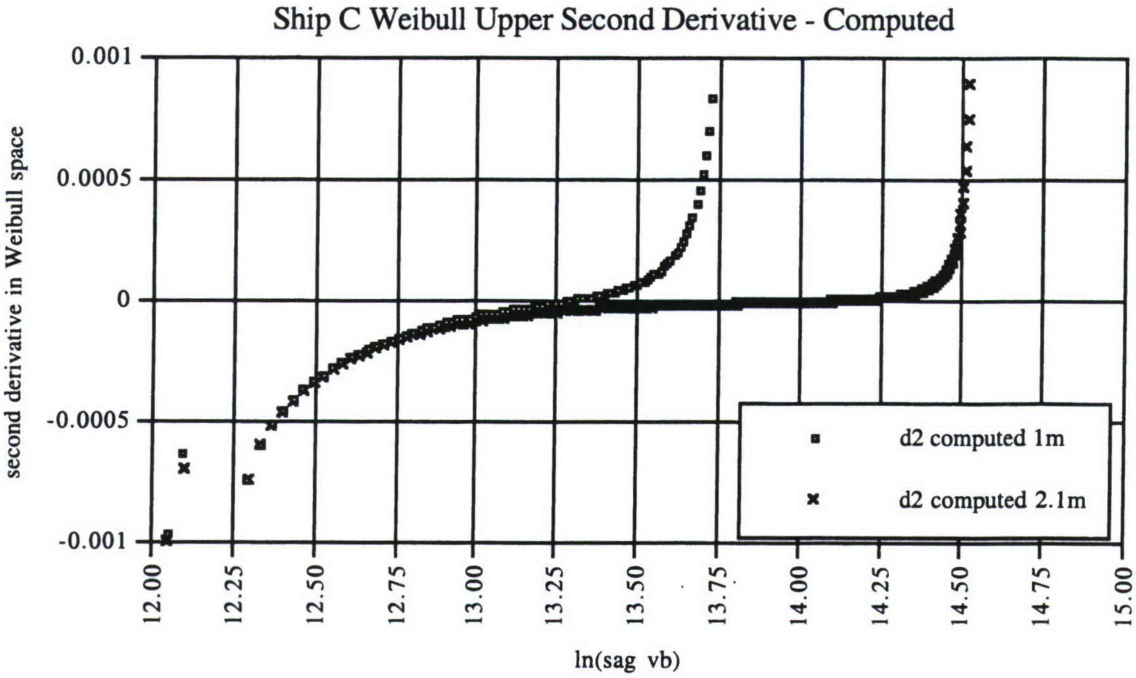
There is a computation and a simulation for each of two different ranges. All fit the larger data values well. No attempt was made to improve the fit at smaller data values since these are not important for design.

The two values of range used were 10^6 and $2.1*10^6$ ft-lt (the Ship C model results were delivered in full-scale units). The initial bin width for computation was 1000 ft-lt / bin. The automatic adjustment for bin size was used in computation mode. The effect of increased range is to push the "turn upward" region further to the right.

8.2.2.1 Largest Usable Value Estimation

As mentioned before, the Ship C fit has a type B (zero function at an infinite boundary) impropriety at the upper end. Since a finite integration range was used and since both computational and simulation results are normalized, the effect is to have rapidly rising probability at the end, hence the abrupt rise at the upper end. Consequently, the full range of either computed or simulated values cannot be used for further computation. The current procedure for estimating the largest useful value is to use the value at which the second derivative of the Weibull plot changes sign.

Plots of the second derivative in Weibull space are shown in Figure 8-3.



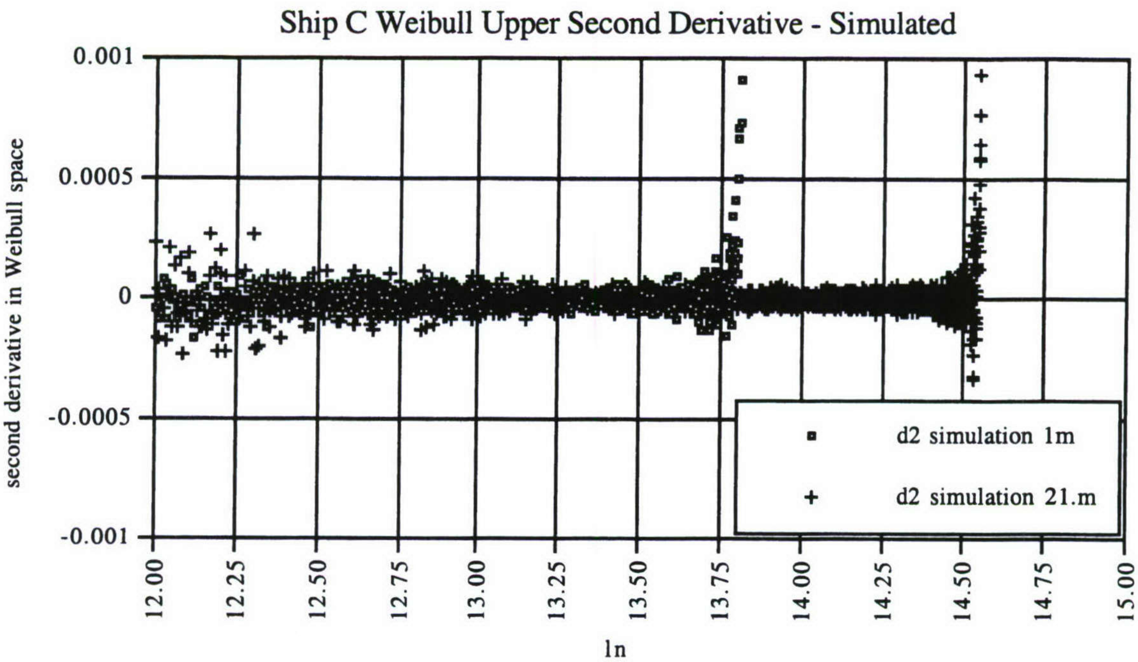


Figure 8-3. Ship C Fit: Second Derivatives By Computation and By Simulation

The randomness in the simulated Weibull space second derivative means that the simulation mode second derivative first crosses zero occurs at too large a $t = lp + hp$ value. Consequently, the $t = lp + hp$ values using computation mode results should be used for estimating the valid region.

9.0 Lifetime Largest (Extreme) Values

Lifetime largest (extreme) values for Ship B, Ship C, and Ship A ship model test runs analyzed in section 6 (Ship A), section 7.2 (Ship B), and section 7.3 (Ship C) are shown in section 9.2 (Ship B), section 9.3 (Ship C), and section 9.4 (Ship A). The arrangement is by increasing complexity of the lifetime estimation techniques used.

The lifetime distribution of the extreme, or largest values, for independent variables identically distributed (the "i.i.d." assumption used here) is given by

$$F = P^n \qquad \text{(3-3b, 3-4b or 4-4b)}$$

where

F	extreme value probability
P	parent or experimental probability
n	lifetime number of independent events

Earlier we expressed the i.i.d. requirements by saying that we will deal with independent events from a statistically stationary process.

There are methods for estimating extreme values when statistically stationary conditions do not exist. They are not employed in this report.

9.1 Extension of Extreme Value Theory to a Combined Load

When both vertical and lateral loads are present we must take into account the combined stress. This combined stress is due to the stress associated with the vertical load acting on the ship cross section plus the stress associated with the lateral load acting on the ship cross section. The stresses, assuming we stay in the linear strain region, are proportional to the sum of (vertical load)/(vertical section modulus) + (lateral load)/(lateral section modulus). We select some point on the cross section where this combination is expected to have its largest value. We may need to look at more than one location on the cross section in order to find the largest combined stress.

Define

$$s = s_{vert} + s_{lat} \qquad (9\text{-}1)$$

where

$$s_{vert} = bending_{vert} / Z_{vert} \qquad (9\text{-}2a)$$

$$s_{lat} = bending_{lat} / Z_{lat} \qquad (9\text{-}2b)$$

$$bending_{vert} = t * \cos(\theta) \qquad (9\text{-}3a)$$

$$bending_{lat} = t * \sin(\theta) \qquad (9\text{-}3b)$$

here

s	total strain (or stress) at the location of interest
s_{vert}	strain (or stress) due to vertical bending
s_{lat}	strain (or stress) due to lateral bending
Z_{vert}	section modulus for vertical bending to the location of interest

Z_{lat} section modulus for lateral bending to the location of interest

t combined load (vertical and lateral bending). This is the value used in the distribution for a particular zone

θ zone angle (=180° for sag with no lateral bending)

Substituting eqs (9-2) and (9-3) into (9-1) we obtain

$$s = \frac{\text{bending}_{vert}}{Z_{vert}} + \frac{\text{bending}_{lat}}{Z_{lat}} = \frac{t * \cos(\theta)}{Z_{vert}} + \frac{t * \sin(\theta)}{Z_{lat}} = t\left(\frac{\cos(\theta)}{Z_{vert}} + \frac{\sin(\theta)}{Z_{lat}}\right) \tag{9-4}$$

Solving for t, we obtain

$$t = \frac{s}{\dfrac{\cos(\theta)}{Z_{vert}} + \dfrac{\sin(\theta)}{Z_{lat}}} \tag{9-5}$$

For θ either 0° or 180°, eq (9-5) becomes $t_{\theta=0,180} = \dfrac{s}{\dfrac{\pm 1}{Z_{vert}}} = s * Z_{vert} = \text{bending}_{vert}$ (9-6a)

For θ either 90° or 270°, eq (9-5) becomes $t_{\theta=90,270} = \dfrac{s}{\dfrac{\pm 1}{Z_{lat}}} = s * Z_{vert} = \text{bending}_{lat}$ (9-6b)

Eq (9-5) does not necessarily represent the worst case since $\cos(\theta)$ and $\sin(\theta)$ are signed quantities. For example, if $\sin(\theta)$ differs in sign from $\cos(\theta)$, look at the other side of the ship - $\sin(\theta)$ will have changed sign. Z_{vert} and Z_{lat} are signed quantities since the "c" in M*c/I may be positive or negative. Consequently, "s" is a signed quantity. However, t must be positive in order to enter our distributions.

We want the most severe condition. This is given by

$$t_k = \frac{|s_j|}{\left|\dfrac{\cos(\theta_k)}{Z_{i,vert}}\right| + \left|\dfrac{\sin(\theta_k)}{Z_{i,lat}}\right|} \tag{9-7}$$

The subscripts have the following meanings: at the i^{th} location on the cross section we set a series of stresses or strains s_j for each of which we evaluate "k" zones.

9.1.1 Cautions When Using Extreme Values

Several cautions are in order when using extreme (largest) values computed from parent distributions.

What credence does the parent (experimental) distribution have? In some (many?) cases it fits the experimental data in the sense that it "looks OK", i.e., passes a visual examination test - the "mark 1" eyeball test. For values to be used for design, statistical tests should be performed on the distribution parameter values estimated from the data no matter the data source (experiment or simulation). We cannot get around the difficulty if we do a simulation: the simulation has (math) model(s) built into it. We may, however, be able to use the results of statistical tests on parameter values to help choose one distribution from several candidate models.

Due to limits on test time and, perhaps, test facility limits, only a relatively small data sample is usually collected. For lifetime loads the distributions of interest are driven by the large magnitude events. This differs from the situation which exists when fatigue loads are being estimated where all cycles of any magnitude are of interest.

Mathematically, the parent distributions used here are unbounded to the right which means they go to infinity. <u>Nothing</u> in the real (physical) world goes to infinity. There are physical limits on how large a bending moment can exist for a particular ship.

Forming an extreme value distribution is an exercise in <u>extrapolation</u>. The major danger in extrapolation is that if the physics changes, the changes in physics will not be reflected in the extreme value distribution. Rare events are apt to be under-represented. One example is the phenomenon of episodic, or rogue waves.

9.2 Ship B

Figure 9-1 shows the combined parent distribution pdf and extreme value distributions for sample sizes of 100, 500, and 1000. The jog in the parent distribution pdf is due to the change in character of the vertical bending high-pass response at low-pass = 16,000 in-lb (16k). "536k" (536,000 in-lb) refers to the maximum range value with pdf > 1.0E-15.

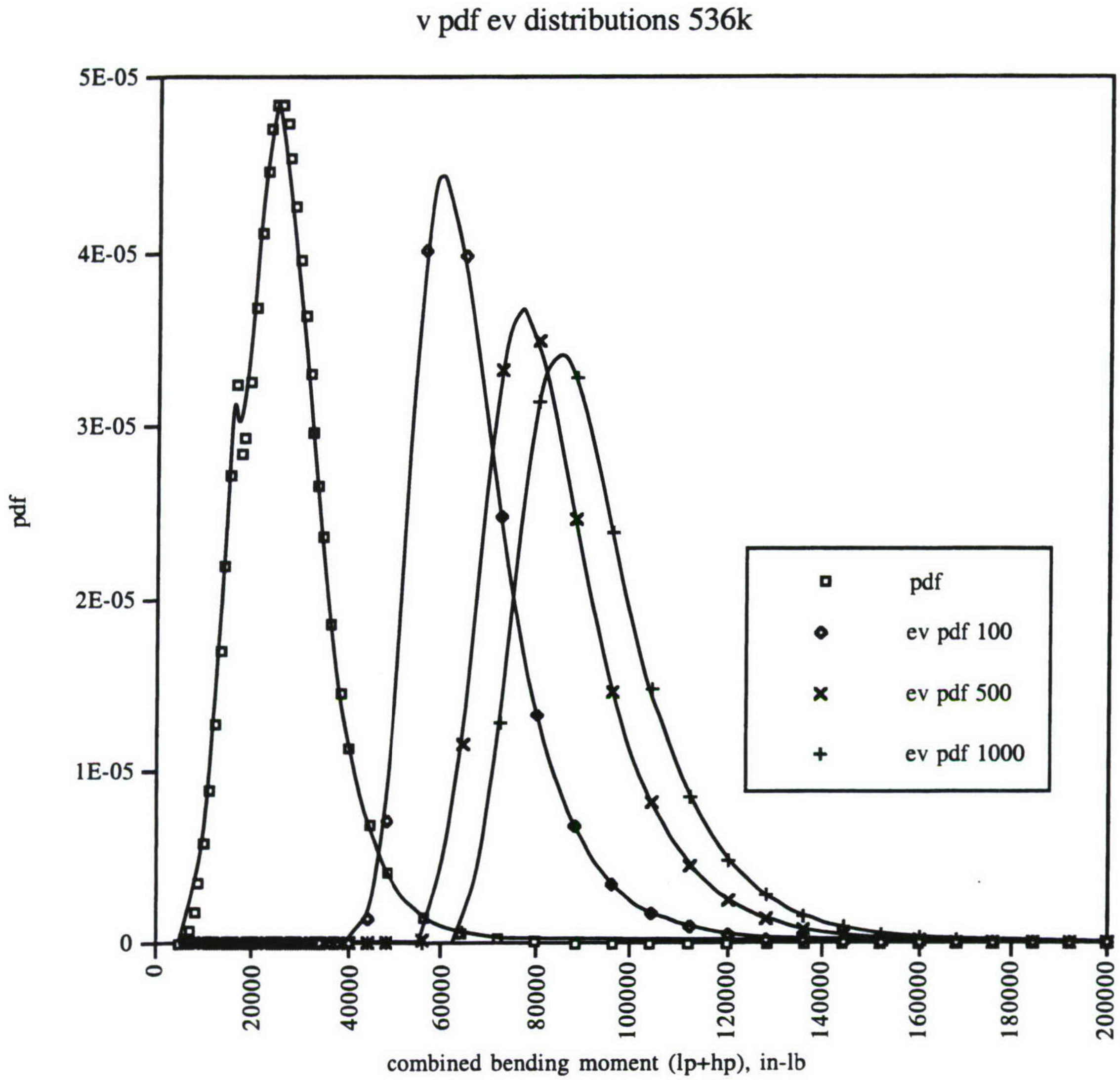


Figure 9-1. Ship B Pdf and Extreme Value Distributions

Figure 9-1 illustrates the application of the technique of obtaining extreme values where the parent distribution has both independent (in this case, low-pass) components and dependent (in this case, high-pass) components. In this case the independent variable has two zones (<16k and >16k). The parameters of the dependent variable have different values for each low-pass zone. The jog at 16k in the parent distribution is due to the change in character of the combined load at 16k.

9.3 Ship C

Figure 9-2 shows the extreme value distributions for sample sizes of 100, 500, and 1000. "2.1m" (2,100,000) refers to the maximum range value used in the computations and simulations. The extreme value curves for sample sizes 500 and 1000 are only partially realized due to the range used for the computations and simulations.

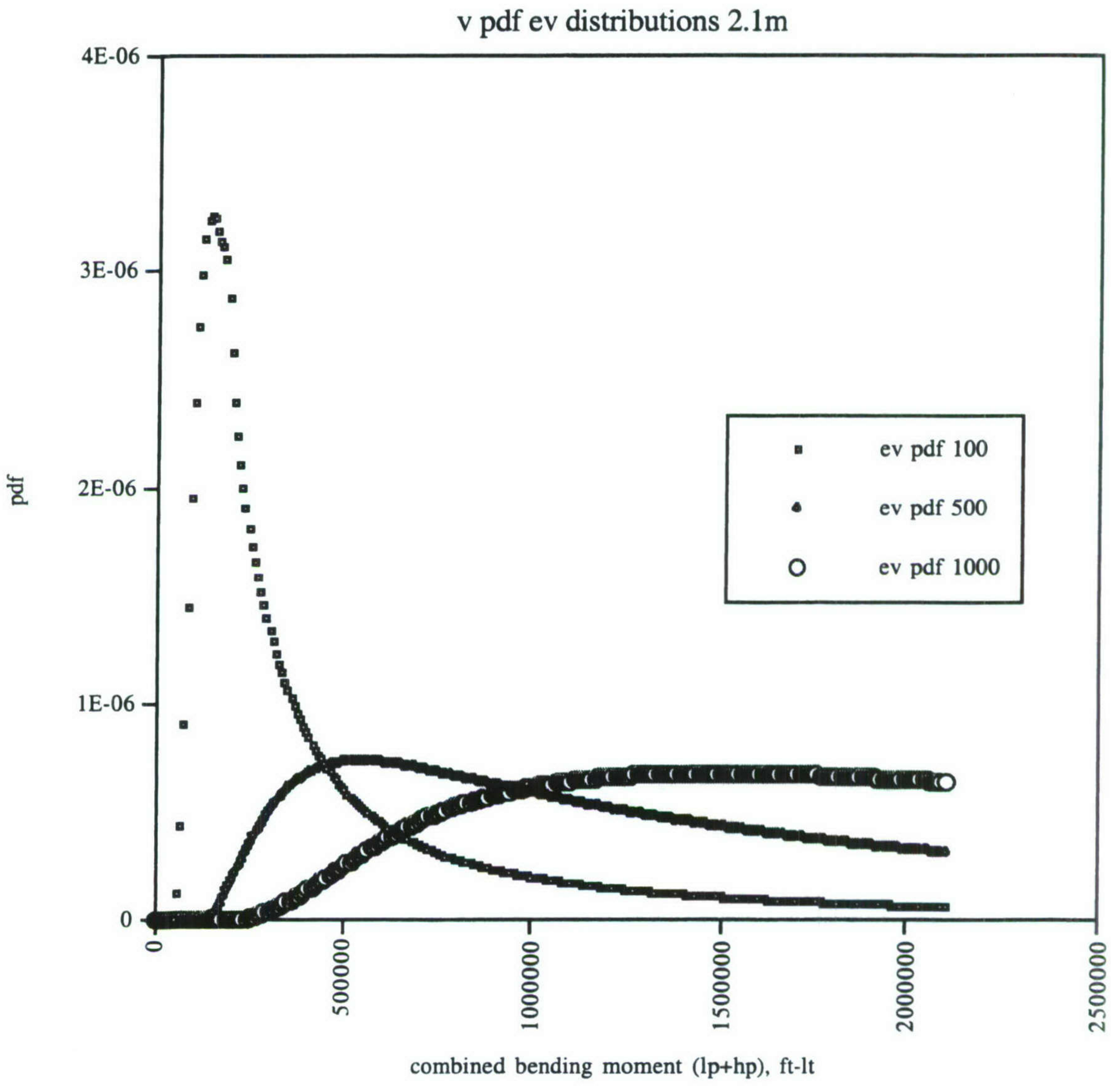


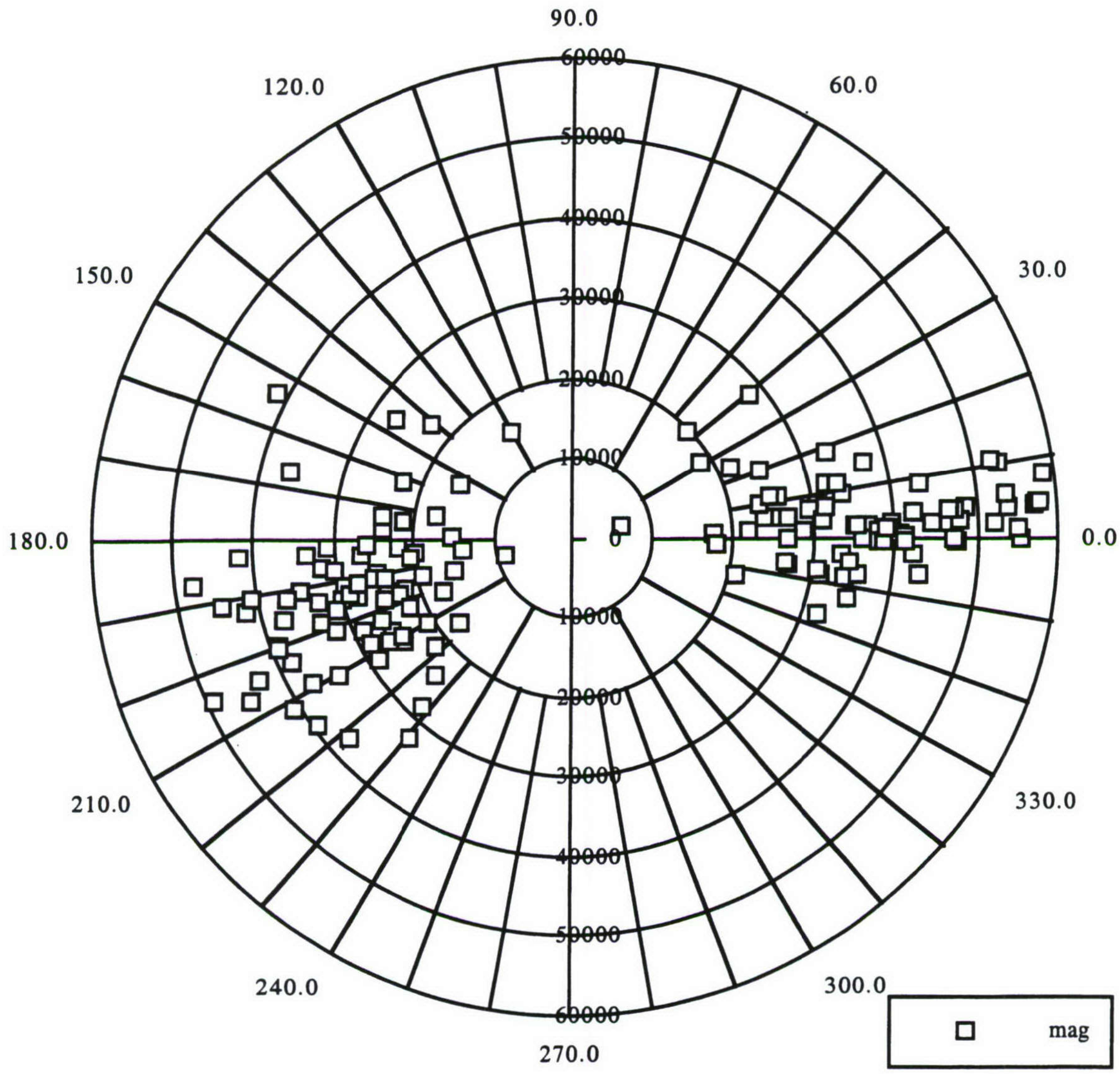
Figure 9-2. Ship C Extreme Value Distributions

Figure 9-2 illustrates another application of the technique of obtaining extreme values where the parent distribution has both independent (in this case, low-pass) components and dependent (in this case, high-pass) components. Here the type of the high-pass distribution differs from the type of the low-pass distribution. For the reasons mentioned in section 9.1.1 realizations of some of these values are not likely or may not be possible.

9.4 Ship A

Both the Ship B and Ship C examples use data from head sea test conditions. The Ship A run selected for analysis has significant values of lateral bending components compared with vertical bending components.

We will estimate the sag load which is associated with 1000 encounters for this condition using the theory developed in section 9.1. The values are from only 2/5 of the run, and are used to illustrate the method. Figure 6-7 is reproduced below. The sag portion (90° to 270°) of the polar plot has 76 events.



(Combined bending moments are in *m-mt*; angles are in *degrees*)
(pure hog = 0 degrees, pure sag = 180 degrees)

Figure 6-7. Combined Vertical and Lateral Bending Magnitude Polar Plot

We divide this arc into zones. The zone size will vary from 59° to 10°. We will treat each such zone as a cell (section 1.2). In each cell, we will then select a distribution and estimate parameters for the distribution using the data in the cell. Finally, we will estimate a load which, when applied to the ship, has a 0.999 chance of not exceeding the stress[58] over the lifetime number of events for all the cells. This example will assume that using a section modulus is valid.

We assume that there are 1000 lifetime events for the condition tested. Table 9-1 shows the number of lifetime events for each zone. We will use the results for the four zones for which parameter estimates have been made. Plots of the data were previously shown in Figure 6-8. A Weibull distribution satisfactorily fit the data in each zone. The "original space" solution values are used and are shown in Table 9-1 (cols 4, 5, and 6).

Table 9-1 shows values which are constant during the calculation. Col 3 is (col 2) * 1000 / sum(col 2), so the entry for sag zone 120-179 = 12 * 1000 / 76 = 157.9 which here is rounded up to 158 since the number of events must always be an integer.

Table 9-1. Lifetime Number of Events and Weibull Parameter Values

1 sag zone	2 data count	3 ltss = 1000 tot	4 x_0	5 sf	6 c
120-179	12	158	not done	not done	not done
180-189	12	158	0	32471	2.8858
190-199	24	316	0	31839	3.4693
200-209	15	197	0	36896	3.5874
210-219	9	118	0	31585	3.2084
220-231	4	53	not done	not done	not done
sum sag =	76	1000			

In Table 9-2 col 2 is the average moment splitting angle associated with the events in a zone. The moment splitting angles for the 12 events in zone 180-189 have an average value = 184.6°. Cols 3 and 4 are the cos and sin of the moment splitting angle.

Cols 5 and 6 are assumed values for the vertical and lateral section moduli (here labeled Z_{xx} and Z_{yy}). The values of Z_{xx} and Z_{yy} are not actual values (knowledge of the actual ship cross section is needed for that), but can be considered as the ratios of the two section moduli. Col 7 is the value of $\frac{\cos(\theta_k)}{Z_{i,\mathrm{vert}}}+\frac{\sin(\theta_k)}{Z_{i,\mathrm{lat}}}$. Col 8 is the value of $\left|\frac{\cos(\theta_k)}{Z_{i,\mathrm{vert}}}\right|+\left|\frac{\sin(\theta_k)}{Z_{i,\mathrm{lat}}}\right|$ which is the required factor [the denominator of eq (9-7)] by which to divide the absolute value of "s" in order to enter the probability distribution for a zone.

[58] Strain is more fundamental. In a linear stress-strain region either may be used.

Table 9-2. Computation for Factor Used in t = s/factor

1 sag zone	2 mean angle	3 cos	4 sin	5 Zxx	6 Zyy	7 cos(th)/Zxx +sin(th)/Zyy	8 factor = |cos(th)/Zxx| +|sin(th)/Zyy|
180-189	184.6	-0.9968	-0.0799	1	1.2	-1.063	1.063
190-199	193.6	-0.9718	-0.2357	1	1.2	-1.168	1.168
200-209	204.3	-0.9115	-0.4113	1	1.2	-1.254	1.254
210-219	212.6	-0.8425	-0.5387	1	1.2	-1.291	1.291

The computation of the lifetime probability of non-exceedance is shown in Table 9-3. The value "s" = 93122 is constant for all zones. The value of the combined load "t" with which to enter the appropriate distribution for a zone is given by s / factor. For zone 180-189 t = 93122 / 1.063 = 87570 [(Table 9-3 col 2) / (Table 9-2 col 8)]. We obtain "P zone" using the Weibull distribution [eq (1-1)] for zone 180-189 as

$$P = 1 - e^{-\left\{\left[\left(t - t_0\right)/sf\right]^c\right\}} = 1 - e^{-\left\{\left[(87570 - 0)/32471\right]^{2.8858}\right\}} = 0.999\,999\,975$$

and "F zone" [eq (3-3b, 3-4b or 4-4b)] for zone 180-189 as

$$F_n = P^n = 0.999\,999\,975^{158} = 0.999\,996\,094$$

Table 9-3. Computation of Lifetime "s" for F = 0.999

1 sag zone	2 s	3 x zone	4 P zone	5 F zone
180-189	93122	87570	0.999 999 975	0.999 996 094
190-199	93122	79711	1.000 000 000	0.999 999 990
200-209	93122	74244	0.999 995 390	0.999 090 458
210-219	93122	72108	0.999 999 272	0.999 913 765
			F =	0.999 000 389

"F" at the bottom of col 4 is the result of multiplying all four "F zone" values together. The value of = 0.999 was our goal: we found a value of "s" (93122), constant for all zones, which results in a probability of non-exceedance of 0.999 when 1000 events are encountered in the operating condition.

9.5 Uncertainty in Extreme Value Estimates

Extreme value estimation is an example of extrapolation. As such, the effect of variability in the parameters of the parent, or experimental distribution, on the computed extreme values is very likely to be magnified (amplified). Sensitivity studies should be performed to quantify this sensitivity. Three examples are shown for estimating the uncertainty in extreme value estimates.

9.5.1 Examples

An expression for the uncertainty is given, as before (section 5.10), by

$$\sigma_{\mathbf{x}}^2 = \mathbf{dvec}^{\mathbf{T}} * \mathbf{pac} * \mathbf{dvec} \qquad (5\text{-}17)$$

where

dvec column vector of derivatives with respect to parameters
pac parameter covariance matrix

Eq (5-17) was derived under the further assumption that we have no bias (eq 5-2).

<u>Extreme Value Uncertainty for Single Cells</u>

Example 1

Cell 200-209 in Table 9-3 has the smallest "F zone" value (0.999 090 458) of any zone. The formula for the extreme value $F = P^n$ for a single cell with a Weibull parent distribution may be solved for x to obtain

$$x = x_0 + sf * \left\{ \ln\left[1 / \left(1 - F^{1/n} \right) \right] \right\}^{1/c} \qquad (B\text{-}12)$$

where "n" is the lifetime number of events.

Using eq (B-12) eq (5-17) becomes

$$\sigma_x^2 = \begin{bmatrix} \partial x/\partial sf & \partial x/\partial c & \partial x/\partial n \end{bmatrix} \begin{bmatrix} \sigma_{sf}^2 & cov(sf,c) & cov(sf,n) \\ cov(c,sf) & \sigma_c^2 & cov(c,n) \\ cov(n,sf) & cov(n,c) & \sigma_n^2 \end{bmatrix} \begin{bmatrix} \partial x/\partial sf \\ \partial x/\partial c \\ \partial x/\partial n \end{bmatrix} \qquad (9\text{-}8)$$

which upon evaluation at x = 74244 becomes

$$\sigma_x^2 = \begin{bmatrix} 2.0122 & -14470.6 & 8.55098 \end{bmatrix} \begin{bmatrix} 7.73157*10^6 & 475.332 & 0 \\ 475.332 & 0.664779 & 0 \\ 0 & 0 & 388.09 \end{bmatrix} \begin{bmatrix} 2.0122 \\ -14470.6 \\ 8.55098 \end{bmatrix} \qquad (9\text{-}8a)$$

leading to $\sigma_x = 11952$. The ratio $\sigma_x / x = 11952/74244 = 0.161$.

The parameter covariance matrix for "sf" and "c" (a 2x2 matrix) forms the upper left hand elements. 388.09 comes from assuming the standard deviation of "n" is 10% [(197/10)^2 = 388.09] where n = 197 is found in Table 9-1 col 3 for zone 200-209. The zero values indicate no interaction between "n" and either "sf" or "c" which is as it should be since "n" is independent of "sf" and "c".

While not immediately obvious, the effect of the variability of the lifetime number of events on the uncertainty is very small. When the standard deviation of the number of lifetime events is increased from 10% to 30% the standard deviation of the uncertainty increases from 11952 to 11962: an increase of 0.08%.

Example 2

Cell 210-219 in Table 9-3 has the next smallest "F zone" value (0.999 913 765) of any zone. Upon evaluation at x = 72108 eq (9-8) becomes

$$\sigma_x^2 = \begin{bmatrix} 2.015479 & -16284.8 & 15.3116 \end{bmatrix} \begin{bmatrix} 1.18407 * 10^7 & 595.125 & 0 \\ 595.125 & 1.12783 & 0 \\ 0 & 0 & 139.24 \end{bmatrix} \begin{bmatrix} 2.015479 \\ -16284.8 \\ 15.3116 \end{bmatrix} \qquad (9\text{-}8b)$$

leading to $\sigma_x = 17673$. The ratio $\sigma_x/x = 17673/74244 = 0.260$.

The parameter covariance matrix for "sf" and "c" (a 2x2 matrix) forms the upper left hand elements. It has been seen before: it is the parameter covariance matrix shown in Table 5-8 col 2.

The variability ratio 0.260 for zone 210-219, having 9 points in its sample, is larger than the variability ratio = 0.161 for zone 200-209 having 12 points in its sample. Not too much should be read into this since the average residual = 511 for zone 200-209 is much smaller than the average residual = 1993 (Table 5-6) for zone 210-219. This difference in average residual is most likely due to sampling variability since the scale factors and slopes for the two zones are comparable (Table 9-1).

<u>Extreme Value Uncertainty for Lifetime Value</u>

We are very interested in the variability in our lifetime estimate s = 93122 (Table 9-3) since this is potentially a value used for ship design. An example of how to proceed is shown in example 3.

Example 3

We will illustrate using the two most severe cells: 200-209 and 210-219. Each cell has a scale factor, a slope, and a number of lifetime events for a total six parameters. The uncertainty eq 5-17 becomes (subscript 1 refers to zone 200-209; subscript 2 refers to zone 210-219)

$$\sigma_s^2 = \begin{bmatrix} \partial s/\partial sf_1 & \partial s/\partial c_1 & \partial s/\partial n_1 & \partial s/\partial sf_2 & \partial s/\partial c_2 & \partial s/\partial n_2 \end{bmatrix} [pac] \begin{bmatrix} \partial s/\partial sf_1 \\ \partial s/\partial c_1 \\ \partial s/\partial n_1 \\ \partial s/\partial sf_2 \\ \partial s/\partial c_2 \\ \partial s/\partial n_2 \end{bmatrix} \tag{9-9}$$

where pac is

$$pac = \begin{bmatrix} \sigma_{sf_1}^2 & cov(sf_1, c_1) & cov(sf_1, n_1) & 0 & 0 & 0 \\ cov(c_1, sf_1) & \sigma_{c_1}^2 & cov(c_1, n_1) & 0 & 0 & 0 \\ cov(n_1, sf_1) & cov(n_1, c_1) & \sigma_{n_1}^2 & 0 & 0 & 0 \\ 0 & 0 & 0 & \sigma_{sf_2}^2 & cov(sf_2, c_2) & cov(sf_2, n_2) \\ 0 & 0 & 0 & cov(c_2, sf_2) & \sigma_{c_2}^2 & cov(c_2, n_2) \\ 0 & 0 & 0 & cov(n_2, sf_2) & cov(n_2, c_2) & \sigma_{n2}^2 \end{bmatrix} \tag{9-9a}$$

The parameter covariance matrix for the combined problem is made up of the parameter covariance matrices for the individual zones. The blocks of zeros show that there is no "cross talk" between the zones. This is consistent with our statement that the cells are independent.

The general form of eq (9-9a), applicable to the entire operational profile response set, is

$$pac = \begin{bmatrix} [pset_1] & [0] & \ldots & [0] \\ [0] & [pset_2] & \ldots & [0] \\ \ldots & \ldots & \ldots & \ldots \\ [0] & [0] & \ldots & [pset_m] \end{bmatrix} \tag{9-10}$$

where

$[pset_1]$	parameter covariance matrix for the first response cell
$[pset_2]$	parameter covariance matrix for the second response cell
$[pset_m]$	parameter covariance matrix for the mth response cell
$[0]$	matrix of zeros

The number of parameters in each "pset" can differ from one response cell to another.

We now turn to obtaining the derivative vector of the parameters. To do this we need the equation for the lifetime value "s" corresponding to a particular lifetime probability of non-exceedance "Pne".

The equation for "s" for two cells (since $x_0 = 0$ for all zones) starts with

$$F = F_1 * F_2 = P_1^{n1} * P_2^{n2} = \left\{1 - e^{-\left[\left(\frac{x_1}{sf_1}\right)^{c1}\right]}\right\}^{n1} * \left\{1 - e^{-\left[\left(\frac{x_2}{sf_2}\right)^{c2}\right]}\right\}^{n2} \qquad (9\text{-}11)$$

From eq (9-7), $x_1 = s / factor_1$ and $x_2 = s / factor_2$, where $factor_1 = \left|\frac{\cos(\theta_1)}{Z_{vert}}\right| + \left|\frac{\sin(\theta_1)}{Z_{lat}}\right|$ and $factor_2 = \left|\frac{\cos(\theta_2)}{Z_{vert}}\right| + \left|\frac{\sin(\theta_2)}{Z_{lat}}\right|$. Substituting into eq (9-11) we have

$$F = \left\{1 - e^{-\left[\left(\frac{(|s|/factor_1)}{sf_1}\right)^{c1}\right]}\right\}^{n1} * \left\{1 - e^{-\left[\left(\frac{|s|/factor_2}{sf_2}\right)^{c2}\right]}\right\}^{n2} \qquad (9\text{-}12)$$

We immediately see a difficulty with eq (9-12) which did not exist when we found the uncertainty for the extreme value for a single cell (example 1 or example 2): eq (9-12) cannot be solved for "s", whereas eq (B-12) is the solution for "x" of $F_1 = P_1{}^{n1}$ when P_1 is a Weibull distribution. This means that the necessary derivatives ($\partial s/\partial sf_1$, $\partial s/\partial c_1$, etc.) cannot be found by direct (partial) differentiation, but must be estimated numerically.

The procedure for doing so (which may be applied to any F composed of many cells each having its own distribution, not necessarily Weibull) is as follows:

1) Select a value of F and find "s". In Table 9-3 F = 0.999 for s = 93122. Call s = 93112 "sb".
2) Perturb a parameter say "sf_1" by some amount $delsf_1$.
3) Solve eq (9-11) for the value of "s" which again makes F = 0.999.
4) The required approximation to $\partial s/\partial sf_1$ is (s - sb) / $delsf_1$.[59]

Proceeding in this manner there is obtained for dvec:

[59] A more precise estimate for the derivatives may be obtained using central differences. The use of central differences is hard to justify given the large amount of variability in the parameter covariance matrix due to sampling variability.

$$\text{dvec} = \begin{bmatrix} 2.3073 \\ -16544 \\ 0 \\ 0.265632 \\ -2088.27 \\ 0 \end{bmatrix}$$

In dvec, the partials $\partial s/\partial n_1$ and $\partial s/\partial n_2$ are numerically approximated by zero.

"pac" is obtained by joining the parameter covariance matrices for the two zones.

$$\text{pac} = \begin{bmatrix} 7.73157*10^6 & 475.332 & 0 & 0 & 0 & 0 \\ 475.332 & 0.664779 & 0 & 0 & 0 & 0 \\ 0 & 0 & 388.09 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1.18407*10^7 & 595.125 & 0 \\ 0 & 0 & 0 & 595.125 & 1.12783 & 0 \\ 0 & 0 & 0 & 0 & 0 & 139.24 \end{bmatrix}$$

Substituting into eq (9-9a), we obtain σ_s = 13853. The ratio σ_s/ s =13853/93122 = 0.149. This ratio is less than either the ratio 0.161 for the zone (200-209) having the smallest "F zone" or the ratio 0.260 for the zone (210-219) having the next smallest "F zone".

While "(two) swallows do not a summer make", the result implies that the uncertainty in the lifetime load may be less than the uncertainty in any of the cells (zones) entering into the computation for the lifetime load. In this case the worst single zone ratio obtained (0.161) might be an upper bound approximation to the lifetime ratio of 0.149.

We have relatively small uncertainties for the examples above. This occurs because the Weibull slope of approximately 3.5 is quite high. *For the same initial sample size, the increase in variability (uncertainty) with decreasing Weibull slope "c" is quite marked.*

10.0 SUMMARY, CONCLUSIONS, AND RECOMMENDATIONS

10.1 Summary

The process in nature in which we are interested (getting bending moments from a seaway and then estimating lifetime loads) is *inherently statistical*. The methods described below are first principle (deal with actual data), are traceable with respect to assumptions and calculations, and the calculations can be computed using current desktop computers (a possible exception might be the situation where a very large amount of simulation is necessary).

The steps (many of which have been known for decades) in the lifetime seaway loads estimation (LSLE) procedure are:

1. Develop an operational profile by assuming the vessel's operating time at-sea can be represented by a collection of cells, i.e., a collection of statistically stationary sea conditions (Table 2-1). The adequacy of this assumption was tested using a 20 year hindcast at a particular location in the North Atlantic Ocean (Table 2-2). The assumption of statistical stationary appeared acceptable.

2. Develop and/or show applicable theory needed for estimating lifetime loads.

   a. Without using order statistics, expressions were developed for estimating the lifetime load for a single cell and for all cells constituting the operational profile.

   b. Using order statistics, expressions were developed for estimating the lifetime load in a cell and for all cells constituting the operational profile. The resulting expression for the lifetime distribution of loads in a cell

$$F_n = P^n \qquad (4\text{-}4b)$$

   is the same using either method ("n" is the number of lifetime events). It does not depend upon the form of the parent or experimental distribution "P". It is derived using the assumption that the events (loads) are statistically independent of each other.

   c. The lifetime distribution of loads for the entire operational profile (the lifetime distribution of loads for all cells) was shown to be given by

$$F(L) = Pne(L) = \prod_{k=1}^{m} [P_k(L)]^{n_k} \qquad (4\text{-}6c)$$

   d. Because the Ship A data analyzed has two bending moment components eq (4-6c) cannot be directly applied. Instead, a method was developed (section 9.1) which, working backwards from a single combined stress (or strain) value at a particular location on the ship's cross section, enables eq (4-6c) to be applied. The result for F = 0.999 is shown in Table 9-3. The key equation is

$$t_k = \frac{|s_j|}{\left|\frac{\cos(\theta_k)}{Z_{i,vert}}\right| + \left|\frac{\sin(\theta_k)}{Z_{i,lat}}\right|} \tag{9-7}$$

3. Develop and/or show applicable theory needed for estimating statistical distribution parameter values.

> *Our goal is to extract as much information as possible from our (usually) one sample of data.*

The use of order statistics facilitates this goal, and is a principal reason for using order statistics.

a. Expressions for the mean, variance and covariance of order statistics are needed to i) compute weights for the data points and so improve the estimation of parameter values and ii) for the estimate of parameter covariance matrices. These expressions were shown along with expressions for the pdf of a single order statistic and for two order statistics (section 4.5).

b. Uncertainties, errors, and sampling distributions applicable to parameter estimation were discussed and defined.

c. The desirability of using of weighted objective functions, where the basic form of the objective function is least squares, was discussed. The weights come from order statistics. An example was presented.

d. Examples using five different fitting methods were shown:
- i) fit in Weibull space, use ln(data) as the independent variable;
- ii) fit in Weibull space, use lnln(1/(1-probability)) as the independent variable;
- iii) fit in original space, use probability as the independent variable, weights = 1;
- iv) fit in original space, use probability as the independent variable, weights = 1/variance found from order statistics;
- v) fit in original space, use probability as the independent variable, weights = inverse of the covariance matrix found from order statistics.

These methods are ranked in order from the least desirable "i)" to the most desirable "v)".

e. A technique for estimating the parameter covariance matrix without doing simulations was shown and illustrated (Tables 5-7 and 5-8).

4. Find a set of independent events. The use of the autocorrelation function was described and illustrated (Figure 6-5). A test showing that the results of using the autocorrelation function were satisfactory was shown (Figure 6-6).

5. Develop a method for representing combined vertical and lateral load events (Figures 6-1 and 6-7) which is compatible with the cell method.

6. Develop methods for representing combined wave and whip distributions (combined low-pass and high-pass distributions) which is compatible with the cell method. Here we deal with high-pass values correlated with low-pass values so that convolution cannot be used to find the sum of the high-pass and low-pass components.

a. For the Ship B data two different sag high-pass distributions were found necessary depending upon the value of the low-pass sag component. Algorithms are summarized in section 7.2.2. Fits using both computation and simulation are shown in Figure 8.1.

b. For the Ship C data two different sag low-pass distributions were found to be necessary depending upon the value of the low-pass component. The high-pass components could not be fitted using Weibull distributions. The high-pass components were successfully fitted using ln normal distributions. Algorithms are summarized in section 7.3.5. Fits using both computation and simulation are shown in Figure 8.2.

7. Estimate lifetime loads (extreme values).

Extreme value distributions for both Ship B and Ship C were found using eq (4-6c) since bending had only one component - in this case, vertical. The results are shown in Figure 9-1 (Ship B) and Figure 9-2 (Ship C). An example of an extreme value calculation for Ship A using eq 9-7, since both vertical and lateral loads are present, is shown in Tables 9-1 through 9-3.

8. Estimate the effect of variability in the parameter estimates. A start on this topic consisted of estimating the parameter covariance matrix (Tables 5-7 and 5-8). Schedule and budget did not permit demonstrating techniques for the interpretation of the parameter estimates, nor of showing the effect of parameter variability on the predicted extreme values. An example of estimating confidence intervals for the three parameter Weibull distribution is shown in [Richardson, 1992].

9. Estimate the uncertainty in statistical distribution values such as bending moments (section 5.10), in an extreme bending moment for a single response cell (section 9.5.1 examples 1 and 2), and for a lifetime extreme bending moment such as might be used for design (section 9.5.1 example 3).

10.2 Conclusions

The primary purpose of this report is to provide a consistent, unified approach, illustrated with examples, for U.S. Navy evaluation of proposed lifetime seaway loads for any surface ship, existing or proposed. The LSLE procedure method(s) presented here are applicable to data analyzed according to first principles, are traceable, and the calculations can be performed on current desktop computers.

The LSLE procedure summarized in section 10.1 for estimating the largest lifetime loads due to a seaway is applicable to any marine vehicle. It applies whether we are doing linear or nonlinear parameter estimation. It does not depend upon any particular statistical distribution - in fact, every response cell may have its own type of distribution, with each such type having a different set of parameter values.

The use of order statistics extracts the most information from our (usually) very limited amount of data. *The smaller the data set, the more important becomes the extraction of as much information as possible.*

The complete method, which includes weighting to obtain some estimate of the estimated parameter variability, does not have to be applied to all data - only to the data which designs the ship.

10.3 Recommendations

The important topic of interpreting the parameter estimates (how "good" are the parameter estimates), and what is the effect of parameter variability on the extreme values did not have examples shown in this report. It is a necessary step needed to give the people who order the ship the engineering knowledge on which to base decisions on the amount of margin to build into the ship. This work should be done.

There are other methods for estimating distribution parameter values such as moment methods. The parameter covariance matrices for these methods should be estimated and compared with the parameter covariance matrices obtained using the methods in this report. We are interested in the methods which give the most reliable estimates. The most reliable method may not give the smallest parameter covariance matrix - recall that Weibull parameter estimation method "e" (using the full data covariance matrix) gave larger parameter covariance matrix estimates than did Weibull parameter estimation method "d" which used only the data variances [Tables 5-8 and 5-9].

The most straightforward, but also the most computationally intensive way to estimate parameter covariances is to use simulation. With simulation an approximation to the parameter distributions, from which the parameter covariances can be computed, can be made. It is of the utmost importance that a good random number generator (rng) be used if simulation is employed. The rng on at least one widely used spreadsheet is not adequate. See Knuth, [1988] for an extensive discussion of rn generators. For results which might be used for ship design rn generators based on two different principles should be used.

Current desktop computers, now available to every engineer, are more powerful than the mainframe computers of a generation ago. One major consequence is that techniques known for years, but previously computationally prohibitive, can now be routinely applied. Some of the techniques described and illustrated in this report are examples of techniques which are not new, but could be, and should be routinely applied. The techniques dealing with parameter uncertainty and uncertainty estimates for derived quantities, such as bending moments, should be applied on a routine basis to the values which are crucial for designing the ship.

11.0 REFERENCES

Abernethy, R. B., Breneman, J. E. , Medlin, C. H. and Reinman, G. L., 1983, *Weibull Analysis Handbook*, AFWAL-TR-83-2079, Aero Propulsion Laboratory [Wright-Patterson AFB].

Bales, Susan, 1983, "Designing Ships to the Natural Environment", Naval Engineers Journal, (March 1983).

Bard, Yonathan, 1974, *Nonlinear Parameter Estimation*, Academic Press, New York.

Burrington, Richarrd S., and May, Donald C., 1970, *Handbook of Probability and Statistics with Tables*, 2nd edition, McGraw-Hill, New York.

Cartwright, D. E. and Longuet-Higgins, M. S., 1956, "The statistical distribution of the maxima of a random function", Proceedings of the Royal Society of London, Series A vol 237 pp 212-232.

Castillo, Enrique, 1988, *Extreme Value Theory in Engineering*, Academic Press, Boston.

David, H. A., *Order Statistics*, 1970, John Wiley & Sons, New York. The theory quoted here is also in David, op. cit. 2nd edition (1982 and David and Nagaraja, op. cit. 3rd edition (2003).

Feller, William, 1968, *An Introduction to Probability Theory and Its Applications*, Vol 1, third edition, John Wiley & Sons, New York.

Gumbel, E. J., 1958, *Statistics of Extremes*, Columbia University Press.

Jasper, N. H., 1956, "Statistical Distribution Patterns of Ocean Waves", Transactions of the Society of Naval Architects and Marine Engineers vol 64.

Kimball, Bradford F., 1960, "On the Choice of Plotting Position on Probability Paper", Journal of the American Statistical Association, vol 55 pp 546-560.

Knuth, Donald E., 1988, *The Art of Computer Programming Vol 2 / Seminumerical Algorithms* 3rd edition, Addison-Wesley, Menlo Park.

[NOC, 1983] *U. S. Navy Hindcast Spectral Ocean Wave Model Climatic Atlas: North Atlantic Ocean*, Naval Oceanography Command Detachment, Asheville, N.C.

Ochi, Michel K., 1964, "Prediction of Occurrence and Severity of Ship Slamming at Sea", Fifth Symposium on Naval Hydrodynamics, Office of Naval Research (ACR-112) pp 545-596.

Psaraftis, Harilaos N., 1976, "Some New Aspects of Slamming Probability Theory", Journal of Ship Research, pp 186-192.

Rice, S. O., 1939, "The Distribution of the Maxima of a Random Curve", American Journal of Mathematics, pp 409-416.

Rice, S.O., 1944, "Mathematical Analysis of Random Noise (Parts I and II)", Bell System Technical Journal, pp 282-332.

Rice, S.O., 1945, "Mathematical Analysis of Random Noise (Parts III and IV)", Bell System Technical Journal, pp 46-156.

Richardson, W. M., 1992, *nlWei3: Estimation of Parameter Values and Parameter Covariances for the Three Parameter Weibull Distribution,* CDNSWC SSPD-93-173-16 October 1992.

Richardson, W. M., 1996, *The Cell Method: a Level III Approach for Estimating Ship Structural Lifetime Reliability*, NSWCCD-TR-65--96/01 March 1996.

Weibull, Waloddi, 1967, Estimation of Distribution Parameters by a Combination of the Best Linear Order Statistic Method and Maximum Likelihood, Technical Report AFML-TR-67-105 [Wright-Patterson AFB].

APPENDIX A: ORDER STATISTICS

Means, Variances, and Covariances

Capital letters such as X and Y refer to actual data values; small letters are used in general formulas. "x" and "y" are variables, "r", "s", and "n" are constants. In particular, "n" refers to the sample size. "r" refers to a particular order number and can range from 1 through n. When considering two order numbers, "s" refers to the second order number and also can range from r+1 through n. "r" is always less than "s".

The expressions below are valid for any continuous parent distribution; analogous expressions exist for the case of a discrete parent distribution.

Means

The k^{th} raw moment is

$$\boldsymbol{\mu}_{r:n}^{(k)} \equiv E\,(X_{r:n}^{k}) \qquad \text{(A-1)}$$

where

$\boldsymbol{\mu}_{r:n}^{(k)}$ — k^{th} raw moment of r in a sample of n.

E — expected value operator.

The bold notation with subscript $_{r:n}$ indicates that eq (A-1) represents a column vector:

$$\boldsymbol{\mu}_{r:n}^{(k)} \equiv \begin{bmatrix} E\,(X_{1:n}^{k}) \\ E\,(X_{2:n}^{k}) \\ \dots \\ E\,(X_{n-1:n}^{k}) \\ E\,(X_{n:n}^{k}) \end{bmatrix}$$

We obtain the mean (average expected value) by setting k = 1 to obtain

$$\boldsymbol{\mu}_{r:n} \equiv E\,(X_{r:n}) \qquad \text{(A-2)}$$

which written as a column vector is

$$\boldsymbol{\mu}_{r:n} \equiv \begin{bmatrix} E\,(X_{1:n}) \\ E\,(X_{2:n}) \\ \dots \\ E\,(X_{n-1:n}) \\ E\,(X_{n:n}) \end{bmatrix}$$

The explicit expression for the mean of the r^{th} in a sample of size n is

$$\boldsymbol{\mu}_{r:n} \equiv E(x_{r:n}) = \int_{-\infty}^{\infty} x * f_r(x) * dx \tag{A-3}$$

where

$f_r(x)$ probability density of the r^{th} size value in a sample of size n (see below for derivation)

Variances and Covariances

The product moments are

$$\boldsymbol{\mu}_{rs:n} \equiv E\,(X_{r:n}X_{s:n}) \tag{A-4}$$

The bold notation with subscripts $_{rs:n}$ indicates that eq (A-4) represents a matrix:

$$\boldsymbol{\mu}_{rs:n} \equiv \begin{bmatrix} E\,(X_{1:n}X_{1:n}) & E\,(X_{1:n}X_{2:n}) & \ldots & E\,(X_{1:n}X_{n-1:n}) & E\,(X_{1:n}X_{n:n}) \\ E\,(X_{2:n}X_{1:n}) & E\,(X_{2:n}X_{2:n}) & \ldots & E\,(X_{2:n}X_{n-1:n}) & E\,(X_{2:n}X_{n:n}) \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ E\,(X_{n-1:n}X_{1:n}) & E\,(X_{n-1:n}X_{2:n}) & \ldots & E\,(X_{n-1:n}X_{n-1:n}) & E\,(X_{n-1:n}X_{n:n}) \\ E\,(X_{n:n}X_{1:n}) & E\,(X_{n:n}X_{2:n}) & \ldots & E\,(X_{n:n}X_{n-1:n}) & E\,(X_{n:n}X_{n:n}) \end{bmatrix} \tag{A-5}$$

We obtain the covariances by replacing the corresponding product moment with

$$\boldsymbol{\sigma}_{rs:n} \equiv E\,(X_{r:n} - \text{mean}_{r:n})\,(X_{s:n} - \text{mean}_{s:n}) \tag{A-6}$$

where

$\boldsymbol{\sigma}_{rs:n}$ covariance of r and s in a sample of size n. When r = s we have the variance $\boldsymbol{\sigma}_{rr:n}$, or $\boldsymbol{\sigma}^2_{r:n}$

$x_{r:n}$, $x_{s:n}$ r^{th} and s^{th} size values in a sample of size n

The result is the covariance matrix eq (A-7):

$$\boldsymbol{\sigma}_{rs:n} \equiv \begin{bmatrix} \text{var}\,(X_{1:n}) & \text{cov}\,(X_{1:n}X_{2:n}) & \ldots & \text{cov}\,(X_{1:n}X_{n-1:n}) & \text{cov}\,(X_{1:n}X_{n:n}) \\ \text{cov}\,(X_{2:n}X_{1:n}) & \text{var}\,(X_{2:n}) & \ldots & \text{cov}\,(X_{2:n}X_{2:n}) & \text{cov}\,(X_{2:n}X_{n:n}) \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \text{cov}\,(X_{n-1:n}X_{1:n}) & \text{cov}\,(X_{n-1:n}X_{2:n}) & \ldots & \text{var}\,(X_{n-1:n}) & \text{cov}\,(X_{n-1:n}X_{n:n}) \\ \text{cov}\,(X_{n:n}X_{1:n}) & \text{cov}\,(X_{n:n}X_{2:n}) & \ldots & \text{cov}\,(X_{n:n}X_{n-1:n}) & \text{var}\,(X_{n:n}) \end{bmatrix}$$

The main diagonal terms (those with both subscripts equal) are the variances, while the other terms are the covariances.

The covariance matrix is symmetric since $\boldsymbol{\sigma}_{rs:n} = \boldsymbol{\sigma}_{sr:n}$.

Explicitly writing out the expression for the variance, we obtain

$$\sigma^2_{r:n} = \int_{-\infty}^{\infty} (x - \text{mean}_{r:n})^2 * f_r(x) * dx \qquad \text{(A-8)}$$

and for the covariance, $r < s$,

$$\sigma_{rs:n} = \int_{-\infty}^{\infty}\int_{-\infty}^{y} (x - \text{mean}_{r:n}) * (y - \text{mean}_{s:n}) * f_{rs}(x,y) * dx * dy \qquad \text{(A-9)}$$

where

$f_{rs}(x,y)$ probability density of the r^{th} and s^{th} size values in a sample of size n (see below for derivation)

Note that if we only are concerned with the variances we may write them as a column vector:

$$\sigma^2_{r:n} = \begin{bmatrix} \int_{-\infty}^{\infty} (x - \text{mean}_{1:n})^2 * f_1(x) * dx \\ \int_{-\infty}^{\infty} (x - \text{mean}_{2:n})^2 * f_2(x) * dx \\ \cdots \\ \int_{-\infty}^{\infty} (x - \text{mean}_{n-1:n})^2 * f_{n-1}(x) * dx \\ \int_{-\infty}^{\infty} (x - \text{mean}_{n:n})^2 * f_n(x) * dx \end{bmatrix} \qquad \text{(A-10)}$$

Probability Densities

To evaluate the above means and covariances requires expressions for $f_r(x)$ and for the joint probability distribution $f_{rs}(x,y)$. One such method, taken from David [1970, §2.1] (which also has other methods), is as follows (his notation using the beta function has been replaced with binomial coefficients):

"... The event $x < X_{(r)} \le x + \Delta x$ may be realized as follows [60]:

$$\frac{r-1 \quad |1| \quad n-r}{x| \quad |x+\Delta x}$$

$X_i \le x$ for $r - 1$ of the X_i, $x < X_i \le x + \Delta x$ for one X_i, and $X_i > x + \Delta x$ for the remaining $n - r$ of the X_i. The number of ways in which the n observations can be so divided into three parcels is

$$\frac{n!}{(r-1)!\,1!\,(n-r)!} = n * C^{n-1}_{r-1} \equiv K_r$$

and each such way has probability

[60] The line indicates a continuum, not division.

$$P^{r-1}(x)\ \ [P(x+\Delta x) - P(x)]\ \ [1 - P(x+\Delta x)]^{n-r}$$

Regarding Δx as small, we have therefore

$$\Pr\{\, x < X_{(r)} \le x + \Delta x \,\} = K_r * P^{r-1}(x)\ p(x)\,\Delta x\ \ [1 - P(x+\Delta x)]^{n-r} + O(\Delta x^2)$$

where $O(\Delta x^2)$ means terms of order $(\Delta x)^2$ and includes the probability of realizations of $x < X_{(r)} \le x + \Delta x$ in which more than one X_i is in $(x, x+\Delta x)$. Dividing both sides by Δx and letting $\Delta x \to 0$, we ... obtain"

$$f_r(x) \;=\; n * C_{r-1}^{n-1} * P^{r-1}(x) * [1 - P(x)]^{n-r} * p(x) \qquad \text{(A-11)}$$

where

$P(x)$	parent probability distribution
$p(x)$	parent probability density distribution

For the joint probability density (again quoting David [1970, §2.2]):

" The joint density function of $X_{(r)}$ and $X_{(s)}$ $(1 \le r < s \le n)$ is conveniently denoted by $f_{rs}(x,y)$. An expression ... may be derived by noting that the compound event $x < X_{(r)} \le x + \Delta x$, $y < X_{(s)} \le y + \Delta y$ is realized (apart from terms having a lower order of probability) by the configuration

$$\begin{array}{ccccc} r-1 & |1| & s-r-1 & |1| & n-s \\ \hline & x|\ \ |x+\Delta x & & y|\ \ |y+\Delta y & \end{array}$$

meaning that $r - 1$ of the observations are less than x, one is in $(x, x + \Delta x)$, etc. It follows that for $x \le y$

$$f_{rs}(x,y) = K_{rs}\, P^{r-1}(x)\, p(x)\, [P(y) - P(x)]^{s-r-1}\, p(y)\, [1 - P(y)]^{n-s}.\text{"} \qquad \text{(A-12)}$$

where

$$K_{rs} \;=\; \frac{n!}{(r-1)!\,(s-r-1)!\,(n-s)!}$$

Here "y" is a dummy variable of integration: the distributions "P(x)" and "P(y)" are the same distribution.

APPENDIX B: WEIBULL DISTRIBUTION TRANSFORMATIONS

The Weibull distribution is described by

$$P = 1 - e^{-\left\{\left[(x-x_0)/sf\right]^c\right\}} \tag{B-1}$$

where

P	=	probability that the value of the variable is x or less
x	=	variable such as load or response
sf	=	scale factor (characteristic value = sf + x_0)
x_0	=	truncation value (value below which the probability is always zero)
c	=	exponent (also referred to as the slope - see below)

Mapping into Weibull Space

Rearrange eq (B-1) to obtain

$$e^{-\left\{\left[(x-x_0)/sf\right]^c\right\}} = 1-P$$

Take natural logs of both sides to obtain

$$-\left\{\left[(x-x_0)/sf\right]^c\right\}\ln(e) = \ln(1-P)$$

Recalling that ln(e) ≡ 1 and multiplying both sides by -1,

$$\left[(x-x_0)/sf\right]^c = -\ln(1-P)$$

Recalling that -ln is equivalent to division,

$$\left[(x-x_0)/sf\right]^c = \ln\left[1/(1-P)\right] \tag{B-2}$$

Taking natural logs again we obtain

$$c * \ln\left[(x-x_0)/sf\right] = \ln\left\{\ln\left[1/(1-P)\right]\right\} \tag{B-3}$$

Now take two points x_2 and x_1. Substituting into eq (B-3) twice we obtain

$$c * \ln\left[(x_2-x_0)/sf\right] = \ln\left\{\ln\left[1/(1-P_2)\right]\right\} \tag{B-4a}$$

and

$$c * \ln\left[\left(x_1 - x_0\right)/sf\right] = \ln\left\{\ln\left[1/\left(1-P_1\right)\right]\right\} \tag{B-4b}$$

Subtracting eq (B-4b) from eq (B-4a) we obtain

$$c * \left\{\ln\left[\left(x_2 - x_0\right)/sf\right] - \ln\left[\left(x_1 - x_0\right)/sf\right]\right\} = \ln\left\{\ln\left[1/\left(1-P_2\right)\right]\right\} - \ln\left\{\ln\left[1/\left(1-P_1\right)\right]\right\}$$

expanding

$$c * \left\{\left[\ln\left(x_2 - x_0\right) - \ln(sf)\right] - \left[\ln\left(x_1 - x_0\right) - \ln(sf)\right]\right\} = \ln\left\{\ln\left[1/\left(1-P_2\right)\right]\right\} - \ln\left\{\ln\left[1/\left(1-P_1\right)\right]\right\}$$

or

$$c * \left[\ln\left(x_2 - x_0\right) - \ln\left(x_1 - x_0\right)\right] = \ln\left\{\ln\left[1/\left(1-P_2\right)\right]\right\} - \ln\left\{\ln\left[1/\left(1-P_1\right)\right]\right\}$$

so "c" in terms of any two distinct points is given by

$$c = \frac{\ln\left\{\ln\left[1/\left(1-P_2\right)\right]\right\} - \ln\left\{\ln\left[1/\left(1-P_1\right)\right]\right\}}{\ln\left(x_2 - x_0\right) - \ln\left(x_1 - x_0\right)} \tag{B-5}$$

This is the equation of the slope of a straight line when we plot $\ln\left(x - x_0\right)$ on the horizontal axis and $\ln\left\{\ln\left[1/\left(1-P\right)\right]\right\}$ on the vertical axis. If we plot $\ln\left\{\ln\left[1/\left(1-P\right)\right]\right\}$ on the horizontal axis and $\ln\left(x - x_0\right)$ on the vertical axis then the slope = 1/c.

Solving for the Scale Factor Using a Weibull Plot

Start with eq (B-3)

$$c * \ln\left[\left(x - x_0\right)/sf\right] = \ln\left\{\ln\left[1/\left(1-P\right)\right]\right\} \tag{B-3}$$

Set the right hand side = 0. At this value the left hand side is

$$c * \ln\left[\left(x - x_0\right)/sf\right] = 0$$

so that

$$\ln\left[\left(x - x_0\right)/sf\right] = 0$$

Take the anti ln to obtain

$$\left(x - x_0\right)/sf = e^0 = 1$$

so

$$sf = \left(x - x_0\right) \tag{B-6}$$

We see that the value of ln(sf) is given by the value of $\ln(x\text{-}x_0)$ when $\ln\left\{\ln\left[1/\left(1-P\right)\right]\right\} = 0$. This means that the value of the scale factor is given by

$$sf = e^{\ln(x-x_0)}\Big|_{\ln\{\ln[1/(1-P)]\}=0} \tag{B-7}$$

Solving for the Scale Factor "sf" When the Fitting Line Equation is Known

If the fitting line in Weibull space of the form $y = c * z + b$ with "c" and "b" are known, then it may be written as

$$\ln\{\ln[1/(1-P)]\} = c * \ln(x - x_0) + b$$

Set the left hand side $\ln\{\ln[1/(1-P)]\} = 0$ so that

$$0 = c * \ln(x - x_0) + b$$

Then

$$\ln(x - x_0) = -b/c$$

Using anti ln we have

$$e^{\ln(x-x_0)} = e^{-b/c} \tag{B-8}$$

In the above subsection we found that $e^{\ln(x-x_0)} = sf$ [eq (B-6)] when $\ln\{\ln[1/(1-P)]\} = 0$ so

$$sf = e^{-b/c} \tag{B-9}$$

Solving for x

Start with eq (B-2)

$$[(x - x_0)/sf]^c = \ln[1/(1-P)] \tag{B-2}$$

Raise both sides to the (1/c) power to obtain

$$(x - x_0)/sf = \{\ln[1/(1-P)]\}^{1/c}$$

Work with the left hand side to obtain

$$x = x_0 + sf * \left(\{\ln[1/(1-P)]\}^{1/c}\right) \tag{B-10}$$

x for an Extreme Value Distribution with a Weibull Parent

The extreme value distribution for any distribution is given by

$$F = P^n \qquad \text{(3-3b) and (4-4b)}$$

If P is a Weibull distribution, then we have

$$F = \left(1 - e^{-\left\{\left[(x-x_0)/sf\right]^c\right\}}\right)^n \tag{B-11}$$

Raising both sides to the (1/n) power we have

$$F^{1/n} = 1 - e^{-\left\{\left[(x-x_0)/sf\right]^c\right\}}$$

Rearranging we have

$$e^{-\left\{\left[(x-x_0)/sf\right]^c\right\}} = 1 - F^{1/n}$$

Taking natural logs

$$-\left\{\left[(x-x_0)/sf\right]^c\right\} = \ln\left(1 - F^{1/n}\right)$$

or

$$\left[(x-x_0)/sf\right]^c = -\ln\left(1 - F^{1/n}\right)$$

Noting that -ln is division

$$\left[(x-x_0)/sf\right]^c = \ln\left[1/\left(1 - F^{1/n}\right)\right]$$

Raising both sides to the 1/c power

$$(x-x_0)/sf = \left\{\ln\left[1/\left(1 - F^{1/n}\right)\right]\right\}^{1/c}$$

and solving for x we have

$$x = x_0 + sf * \left\{\ln\left[1/\left(1 - F^{1/n}\right)\right]\right\}^{1/c} \tag{B-12}$$

Note that this the same formula as (B-10) except that P is replaced with $F^{1/n}$.

APPENDIX C: ESTIMATE FOR PARAMETER COVARIANCE MATRIX

The derivation of an estimate for the parameter covariance matrix follows. We proceed according to Bard [1974] (asterisks * indicate estimated values):

"... At the minimum [of the objective function] we have

$$\partial\Phi(\theta^*, W) / \partial\theta = 0 \tag{7-5-1}$$

[where **W** represents the measured data values]

Suppose we varied the data slightly, replacing $\mathbf{w}$ by $\mathbf{w} + \delta\mathbf{w}$. This would cause our minimum to shift from θ^* to $\theta^* + \delta\theta^*$, where we must have

$$\partial\Phi(\theta^* + \delta\theta^*, W + \delta W) / \delta\theta = 0 \tag{7-5-2}$$

Expanding Eq (2) in Taylor's series and retaining only terms up to first order, we find after subtracting Eq (1)

$$\left(\delta^2\Phi / \delta\theta^2\right)\delta\theta^* + \left(\partial^2\Phi / \partial\theta\,\partial\mathbf{w}\right)\delta\mathbf{w} \approx 0 \tag{7-5-3}$$

so that approximately

$$\delta\theta^* \approx -\mathbf{H}^{*^{-1}}\left(\partial^2\Phi / \partial\theta\,\partial\mathbf{w}\right)\delta\mathbf{w} \tag{7-5-4}$$

where

$$\mathbf{H}^* = \partial^2\Phi / \partial\theta^2\Big|_{\theta=\theta^*}$$

The desired covariance matrix $\mathbf{V}_\theta$ is defined by

$$\mathbf{V}_\theta \equiv \mathbf{E}\left(\delta\theta^*\,\delta\theta^{*^T}\right) \tag{7-5-5}$$

so that

$$\mathbf{V}_\theta \approx \mathbf{E}\left(\mathbf{H}^{*^{-1}}\left(\partial^2\Phi / \partial\theta\,\partial\mathbf{w}\right)\delta\mathbf{w}\,\delta\mathbf{w}^\mathbf{T}\left(\partial^2\Phi / \partial\theta\,\partial\mathbf{w}\right)^\mathbf{T}\mathbf{H}^{*^{-1}}\right) \tag{7-5-6}$$

The quantities $\mathbf{H}^*$ and $\partial^2\Phi / \partial\theta\,\partial\mathbf{w}$ are evaluated at $\theta=\theta^*$ and at the actual sample **W**. Hence they are constants, and can be taken outside the expectation sign in Eq (6)

$$\mathbf{V}_\theta \approx \mathbf{H}^{*^{-1}}\left(\partial^2\Phi / \partial\theta\,\partial\mathbf{w}\right)\mathbf{V}_\mathbf{w}\left(\partial^2\Phi / \partial\theta\,\partial\mathbf{w}\right)^\mathbf{T}\mathbf{H}^{*^{-1}} \tag{7-5-7}$$

where $\mathbf{V}_\mathbf{w}$ is the covariance matrix of the data, i.e.,

$$\mathbf{V_w} \equiv \mathbf{E}\left(\boldsymbol{\delta}\mathbf{w} \ \ \boldsymbol{\delta}\mathbf{w}^T \right) \tag{7-5-8}$$

... "

All terms in Eq (7-5-7) are known except for the covariance matrix of the data $\mathbf{V_w}$. $\mathbf{V_w}$ is the actual covariance matrix of the data. To compute this experimentally with no sampling error would require all of the (infinite number of) samples which we have not yet drawn. Put another way, the experimental data covariance matrix has its own sampling distribution.

NAVSEA
CARDEROCK
Surface Warfare Center Division